THE LANCET
Psychiatry

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Taquet M, Luciano S, GeddesJR, Harrison PJ. Bidirectional associations between COVID-19 and psychiatric disorder: retrospective cohort studies of 62 354 COVID-19 cases in the USA. *Lancet Psychiatry* 2020; published online Nov 9. https://doi.org/10.1016/S2215-0366(20)30462-4.

Supplementary Methods

TriNetX network – Data

In this section, we provide more details about the TriNetX network, the data source, its advantages and disadvantages.

The data is stored onboard a TriNetX appliance – a physical server residing at the institution's data centre or a virtual hosted appliance. The TriNetX platform is a fleet of these appliances connected into a federated network able to broadcast queries to each appliance. Results are subsequently collected and aggregated.

Once the data is sent to the network, it is mapped to a standard and controlled set of clinical terminologies and undergoes a data quality assessment including 'data cleaning' that rejects records which do not meet the TriNetX quality standards (see next section). HIPAA compliance of the clinical patient data is achieved using de-identification. Different data modalities are available in the network. They include demographics (coded to HL7 version 3 administrative standards), diagnoses (represented by ICD-10-CM codes), procedures (coded in ICD-10-PCS or CPT), and measurements (coded to LOINC). While extensive information is provided about patients' diagnoses and procedures, other variables (such as socioeconomic and lifetime factors are not comprehensively represented).

The advantage of EHR data over insurance claim data is that both insured and uninsured patients are included. An advantage of EHR data over survey data is that they represent the diagnostic rates in the population presenting to healthcare facilities. This provides an accurate account of the burden of specific diagnoses on healthcare systems. The downside of relying on diagnoses is that they obviously do not account for undiagnosed patients who might be suffering from the illness but did not seek medical attention. A general limitation of EHR data is that a patient may be seen in different HCOs for different parts of their care and in one HCO is not part of the federated network then part of their medical records may not be available. Using a network of HCOs (rather than a single HCO) limits this possibility but does not fully remove it.

TriNetX network – Quality control

Data quality assessment followed the strategy developed by Kahn et al.[1] where the data are reviewed for conformance (adherence to specified standards and formats), completeness (quantifying data presence or absence) and plausibility (believability of the data from a clinical perspective). There are pre-defined metrics for each of the above assessment categories. Results for these metrics are visualised and reviewed for each new site that joins the network as well as on an ongoing basis. Any identified issue is communicated to the data provider and resolved before continuing data collection.

The basic formatting of contributed data is also checked (e.g. to ensure that dates are properly represented). Records are checked against a list of required fields (e.g., patient identifier) and rejects those records for which the required information is missing. Referential integrity checking is done to ensure that data spanning multiple database tables can be successfully joined together. As the data are refreshed, changes in volume of data over time is monitored to ensure data validity. At least one non-demographic fact for each patient is required for them to be counted in the dataset. Patient records with only demographics information are discarded.

The software also undergoes quality control. The engineers testing the software are independent from the engineers developing it. Each test code is checked by two independent testing engineers. Each piece of software is tested extensively against a range of synthetic data (i.e. generated for the purpose of testing) for which the expected output is established independently. If the software fails to return this output, then the software is deemed to have failed the test and is examined and modified accordingly. For statistical software (including that used for propensity score matching, for Kaplan-Meier analysis, etc), an additional quality control step is implemented. Two independent codes are written in two different programming languages (typically R and python) and the statistical results are compared.

If discrepancies are identified, then the codes are deemed to have failed the test and are examined and modified accordingly. All the code is reviewed independently by another engineer.

The test strategy follows three levels of granularity:

1. Unit tests: These test specific blocks, or units, of code that perform specific actions (e.g. querying the database).
2. Integration tests: These ensure that different components are working together correctly.
3. End-to-end tests: These tests run the entire system and check the final output.

Definition and encoding of covariates

Covariates for the analysis of psychiatric antecedents (i.e. testing whether psychiatric illness is a risk factor for a diagnosis of COVID-19) included established and suspected risk factors for COVID-19. In the language of causal diagrams, controlling for all risk factors for COVID-19 would achieve the back-door criterion allowing estimation of the causal effect of a diagnosis of COVID-19 onto subsequent psychiatric illness.[2] The following risk factors were therefore used (with the ICD-10/CDC code in brackets when indicated):

1) **Age**
2) **Sex**
3) **Race** encoded as 6 separate dichotomous variables: White (2106-3), Black or African American (2054-5), American Indian or Alaska Native (1002-5), Asian (2028-9), Native Hawaiian or Other Pacific Islander (2076-8), or Unknown Race (2131-1)
4) **Obesity** encoded as one dichotomous variable and one categorical variable: Overweight and obesity (E66) and body mass index (categorised into < 25 kg/m^2, 25-30 kg/m^2, ≥ 30 kg/m^2 corresponding to the WHO definition of normal weight, overweight, and obesity respectively).
5) **Hypertension** encoded as 2 dichotomous and 2 categorical variables: Hypertensive diseases (I10-I16), the now deprecated version that was used until 2018 Hypertension diseases (I10-I15), measurements of systolic blood pressure (categorised into $<$ 140mmHg, 140-160mmHg, and $\geq$ 160mmHg), and diastolic blood pressure (categorised into $<$ 90mmHg, 90-100mmHg, and $\geq$ 100mmHg). These thresholds correspond to normal blood pressure, Grade 1, and Grade 2 hypertension according to the 2020 International Society of Hypertension Global Hypertension Practice Guidelines.
6) **Diabetes mellitus** encoded as 2 dichotomous variables: Type 1 diabetes mellitus (E10) and Type 2 diabetes mellitus (E11)
7) **Chronic lower respiratory diseases** encoded by each sub-category of the corresponding ICD-10 group (so matching is achieved for each individual subcategory): Bronchitis, not specified as acute or chronic (J40), Simple and mucopurulent chronic bronchitis (J41), Unspecified chronic bronchitis (J42), Emphysema (J43), Other chronic obstructive pulmonary disease (J44), Asthma (J45), Bronchiectasis (J47)
8) **Nicotine dependence** encoded as the corresponding ICD 10 diagnosis (F17.2)
9) **Heart diseases** encoded as 2 categorical variables: Ischaemic heart disease (I20-I25) and Other forms of heart disease (I30-I52)
10) **Chronic kidney disease** encoded as 2 dichotomous variables: Chronic kidney disease (N18) and Hypertensive chronic kidney disease (I12)

For the analysis of the psychiatric sequelae of COVID-19, we add to the set of covariates risk factors for a more severe COVID-19 illness. As risk factors for severity, we used the established factors associated with COVID-19 death based on a large community sample[3], which include (besides those already included in the list above):

11) **Chronic liver disease** encoded as 8 categorical variables: Alcoholic liver disease (K70), Hepatic failure, not elsewhere classified (K72), Chronic hepatitis, not elsewhere classified (K73), Fibrosis and cirrhosis of liver (K74), Fatty (change of) liver, not elsewhere classified (K76.0), Chronic passive congestion of liver (K76.1), Portal hypertension (K76.6), Other specified diseases of liver (K76.8)
12) **Stroke** encoded as the dichotomous variable Cerebral infarction (I63)
13) **Dementia** encoded as 4 dichotomous variables: Vascular dementia (F01), Dementia in other diseases classified elsewhere (F02), Unspecified dementia (F03), and Alzheimer's disease (G30)

14) **Cancer and haematological cancer in particular** encoded as 2 dichotomous variables: Neoplasms (C00-D49) and Malignant neoplasms of lymphoid, hematopoietic and related tissue (C81-C96)
15) **Organ transplant** encoded as 2 dichotomous variables: Renal Transplantation Procedures and Liver Transplantation Procedures
16) **Rheumatoid arthritis** encoded as 2 dichotomous variables: Rheumatoid arthritis with rheumatoid factor (M05) and Other rheumatoid arthritis (M06)
17) **Lupus** encoded as a dichotomous variable corresponding ICD-10 code (M32)
18) **Psoriasis** encoded as a dichotomous variable corresponding ICD-10 code (L40)
19) **Other immunosuppression** encoded as a dichotomous variable "Certain disorders involving the immune mechanism" (D80-D89)

Note for all diagnostic categories, an individual was considered positive if the diagnostic was recorded at least once in their health record. For measurement variables represented as categorical variables (i.e. BMI and blood pressures), all available measurements for all individuals were used so that propensity score matching sought to define cohorts with similar numbers of measurements falling into each category.

Definition of control cohorts

The six control cohorts used for the analysis of psychiatric sequelae were defined based on their having had a diagnosis of an acute health event on or after January 20, 2020, namely (with the ICD-10 code in brackets). We selected 6 health events which represent a broad range of common presentations (some with clinical presentations similar to COVID-19 and some with very different presentations). This set of control health events was predefined and no other health event was tested. They are:

1) **Influenza cohort** – any of the following diagnoses: Influenza due to certain identified influenza viruses (J09), Influenza due to other identified influenza virus (J10), Influenza due to unidentified influenza virus (J11). This health event was selected as the most clinically similar presentation to COVID-19.
2) **Other respiratory tract infection cohort** – any of the following diagnoses: acute upper respiratory tract infection (J00-J06), influenza and pneumonia (J09-J18), other acute lower respiratory tract infections (J20-J22) and excluding all those who had a diagnosis of COVID-19 at any point on or after January 20, 2020. This health event was selected as another health event with a similar clinical presentation to COVID-19 and with an incidence in the period from January 20, 2020 to July 31, 2020 that was higher than influenza and, in particular, with a reasonably high incidence after April 1, 2020 (given us the ability to test the "context" hypothesis).
3) **Skin infection cohort** – Infection of the skin and subcutaneous tissue (L00-L08). This health event was selected as a common infective event that does not affect the respiratory tract.
4) **Cholelithiasis cohort** – Cholelithiasis (K80). This health event and the next one were selected as common surgical presentations.
5) **Urolithiasis cohort** – Urolithiasis (N20-N23)
6) **Fracture of a large bone cohort** – any of the following diagnoses: Fracture of lumbar spine and pelvis (S32), Fracture of shoulder and upper arm (S42), Fracture of forearm (S52), Fracture of femur (S72), Fracture of lower leg including ankle (S82). This health event was selected as it is largely independent of prior medical conditions, and almost always leads to presentation for treatment.

Definition of outcomes for the analysis of psychiatric sequelae

All outcomes for the analysis of psychiatric sequelae were defined as a diagnosis recorded between 14 days and 90 days of the diagnosis of COVID-19 (or control health event). The exclusion of the first 14 days from this window limits the contamination of the samples with late recordings in the patient's EHR of pre-existing diagnoses (e.g. a patient admitted with COVID-19 who has a pre-existing mood disorder which is recorded in their EHR on the first day of admission) and the potential for misdiagnosis during acute illness, e.g. between delirium and dementia; the time to recover from COVID-19 for inpatients is about 2 weeks[4] and the Centers for Disease Control (CDC) recommend that outpatients self-isolate for 10 days after the onset of symptoms[5].

The following diagnostic codes were used to define secondary outcomes in the analysis of psychiatric sequelae:

1) Psychotic disorder: F20-F29
2) Mood disorder: F30-F39
    a) Mania/Bipolar: F30-F31
    b) Depressive episode: F32
3) Anxiety disorder: F40-F48
    a) Phobia: F40
    b) Other anxiety disorder: F41
        i) Panic disorder: F41.0
        ii) Generalized anxiety disorder: F41.1
    c) Obsessive compulsive disorder (OCD): F42
    d) Reaction to severe stress: F43
        i) Acute stress reaction: F43.0
        ii) Post-traumatic stress disorder: F43.1
        iii) Adjustment disorder: F43.2
    e) Dissociative disorder: F44
    f) Somatoform disorder: F45
    g) Other neurotic disorder: F48
4) Insomnia: either Nonorganic Insomnia (F51.0) or Insomnia (G47.0)
5) Dementia: any of Vascular Dementia (F01), Dementia in Other Diseases Classified Elsewhere (F02), Unspecified Dementia (F03), Alzheimer's disease (G30), Frontotemporal dementia (G31.0), or Dementia with Lewy bodies (G31.83)

In addition, to calculate the probability of developing insomnia without a concurrent diagnosis of anxiety, we defined an additional outcome "insomnia OR anxiety" which is considered met if either insomnia or anxiety disorder is met (no other Boolean operators are allowed in the definition of outcomes within the TriNetX platform). This allowed us to calculate the probability of having insomnia but not anxiety within the 14 days–90 days period as follows: P(Insomnia, NOT Anxiety) = P(Insomnia OR Anxiety) - P(Anxiety) which are the results presented in Fig. 3 (appendix p. 27).

Sensitivity analyses

We conducted additional analyses to (i) test the sensitivity of the findings to cohort and outcome specifications, and to (ii) explore possible explanations for the findings. Here we describe the rationale and design of these additional analyses.

For both the analysis of the psychiatric sequelae and antecedents of COVID-19, sensitivity analyses were conducted by focusing on confirmed cases of COVID-19, by excluding participants with unknown race, and by adjusting for socioeconomic factors as follows:

Race known: In the primary analysis, we elected to include individuals whose race was unknown. This is because we wanted to achieve the most representative sample of the general population, and a significant fraction of the population has no recorded race in their EHR (Table 1 in the manuscript). We used "unknown race" as a variable on which cohorts must be matched to guarantee that similar numbers of individuals with unknown race would be present in both cohorts. So long as the race distribution among the "unknown race" group is similar between the two cohorts, including them does not bias the findings. In these additional analyses, we tested this by excluding all individuals with unknown race from the cohorts and conducting the same analytic approach as the primary analyses.

Adjusting for socioeconomic factors: Socioeconomic factors are potentially associated with both COVID-19 and psychiatric illness. A limitation of EHR studies is that such factors are not comprehensively recorded. However, indicators of socioeconomic deprivation are encoded under the category Z59 ("Problems related to housing and economic circumstances") of ICD-10 which comprises homelessness, inadequate housing, and extreme poverty among other things and which was present in about 250,000 patients of the TriNetX network. In this sensitivity analysis, we included Z59 as a covariate. It is also worth noting that socioeconomic status is largely controlled for

by matching for race and BMI, which applies to all our cohorts. For instance, a recent study of risk factors for COVID-19 has shown that accounting for age, sex, and Black Asian and Minority Ethnicity (BAME), the deprivation score has an odd-ratio of merely 1.03 whereas accounting for age, sex, and deprivation, BAME has an odd ratio of 1.67[6].

Confirmed COVID-19: To capture the early stage of the pandemic (during which the newly developed WHO codes were probably not yet in routine use, the definition of COVID-19 as an outcome (in the analysis of psychiatric antecedents) and as a cohort (in the analysis of psychiatric sequelae) encompassed: COVID-19 with virus identified (U07.1), COVID-19 without virus identified (U07.2), Pneumonia due to SARS-associated coronavirus (J12.81), Other coronavirus as the cause of disease classified elsewhere (B97.28), and Coronavirus infection unspecified (B34.2). In this sensitivity analysis, we analysed the differential incidence of confirmed COVID-19 only (code U07.1) between the two cohorts. Similarly, in a sensitivity analysis of the psychiatric sequelae, we defined the main cohort as having had confirmed COVID-19 and compared it to the 6 control cohorts.

For the analysis of the psychiatric sequelae of COVID-19, we conducted another four sensitivity analyses and two additional analyses to test for factors that might explain the finding that COVID-19 leads to significantly higher rates of psychiatric sequelae than the control health events used. The additional sensitivity analyses included:

RNA/Antigen-confirmed COVID-19: COVID-19 diagnosis is typically confirmed with either RNA/Antigen test or with antibody test. The former is usually performed during the early symptomatic phase of the illness while the latter can be performed at a later stage. As a result, the timing of the diagnosis with respect to the timing of psychiatric sequelae might be heterogeneous in the population. In this sensitivity analysis, we restricted patients in the COVID-19 cohort to those with an RNA/Antigen-confirmed test.

Comparing to sequelae before the pandemic: In the primary analysis, we compared the rate of psychiatric sequelae between COVID-19 and other health events within the period from January 20 to August 1, 2020. This aimed to compare health events which all occur within the same time window. However, this was a very unusual context which might impact the rate of psychiatric sequelae of any health event. So long as the 'pandemic' context impacts the rates of psychiatric sequelae of the different health events in a similar way, the interpretation of the comparison between the groups remains straightforward. But the interpretation of the findings would become more convoluted if the 'pandemic' context increased the apparent rate of psychiatric sequelae of COVID-19 and decreased that of other health events. For instance, it might be that having a bone fracture is less psychologically traumatic during the pandemic because opportunities for physical activities are already restricted. To address this issue, in this sensitivity analysis, we compared the rate of psychiatric sequelae of COVID-19 (diagnosed between January 20 and August 1, 2020) to those of the other 6 health events in a 'normal' context (taken to be the same time window – Jan 20 to Aug 1 – in 2019 rather than 2020).

Unmatched cohorts: Matching the cohorts for confounding factors was a key component of the primary and sensitivity analyses. One limitation of matching approaches is that the resulting cohorts are necessarily more restricted than the original ones and so is the external validity of the findings. Comparing unmatched cohorts allows the outcomes to be compared between the *whole* cohorts. The necessary downside is that any association observed might be due to the health event (e.g. COVID-19 vs influenza) or the baseline confounding variables (e.g. Type 2 diabetes) or both. Unmatched analysis thus provides complementary findings whose interpretation needs to be informed by careful examination of differences in baseline characteristics.

Patients with a subsequent health visit: In the primary analysis, we included all patients regardless of whether they had an encounter with a healthcare organization. This was deemed important because a patient who has had no encounter throughout the follow-up period (14 days–90 days after the health event) but then has an encounter after that period is recorded as not having had the outcome rather than counted as "lost to follow-up" which might artificially inflate the outcome rate. However, because the proportion of such patients might differ between cohorts, we repeated the analysis while focusing only on patients with at least one encounter with a healthcare organization between 14 days and 90 days after their health events.

The two analyses used to test for factors that might explain the association between COVID-19 and subsequent psychiatric diagnosis were:

Severity of the illness: One possibility is that COVID-19 leads on average to a more severe illness and that it is the severity of the illness which leads to higher rates of psychiatric sequelae rather than the COVID-19 illness itself. We tested this hypothesis in two steps. First, we tested whether a more severe COVID-19 illness indeed leads to higher incidence of psychiatric sequelae than a less severe presentation. This is achieved by comparing the incidence of psychiatric sequelae (coded as a diagnosis of F20-F48) among patients with COVID-19 who had to be admitted as inpatient to that among those who remained outpatients. The inpatient status was encoded by the presence of any of the following in the patient's EHR from 4 days before the diagnosis to 2 weeks after: Inpatient Visit, Short Stay Visit, Critical Care Services, Hospital Inpatient Services, Initial Inpatient Consultation Services. Second, we tested whether the difference in incidence of psychiatric sequelae remains when we limit the cohorts to those who were outpatients (e.g. comparing the psychiatric sequelae among outpatients with COVID-19 to that among outpatients with influenza). If the findings can be entirely explained by differences in the severity of the illness, then no significant difference in psychiatric sequelae would be observed once the cohorts are limited to outpatients.

Contextual factors: A second possible explanation is that differences in the clinical context might have driven differences in the rate of psychiatric sequelae between COVID-19 and other control events. All cohorts were defined so that the health event occurred between January 20, 2020 and August 1, 2020 which limits the possibility for contextual differences. However, within the period from January 20, 2020 to August 1, 2020, there remains differences in the timing of health events between cohorts. A clear example is that respiratory tract infections were mostly diagnosed at the beginning of that period whereas COVID-19 was mostly diagnosed in the second half of that period. It might follow that patients in the control cohorts had their health event in a different – and perhaps simpler – clinical context (e.g. less saturated healthcare services, absence of self-isolation rules for those presenting symptoms of COVID-19, less fear associated with hospital visits, etc). However different the clinical context might be, it is possible that such differences have impacted the incidence of psychiatric sequelae among people who experienced an acute health event within this context. To test this, we first compared cohorts with the same index event diagnosed after April 1 vs between January 20 and March 31, to assess whether the clinical context indeed had an impact on psychiatric sequelae. This was done for COVID-19 as well as all control health events except for influenza (as the incidence of influenza after April 1 was too low for meaningful inference–see Fig. 22 in this appendix p. 45). The date of April 1 was chosen because it roughly corresponds to the first peak of the pandemic in the USA when many states started imposing stay-at-home orders, and because it allowed us to keep sufficient number of subjects in both the cohort before April 1 and the cohort after April 1 for all health events (apart from influenza).

Second, we repeated the primary analysis (i.e. comparing the psychiatric sequelae between COVID-19 and other control index events) while restricting the cohorts to those having had the index event on or after April 1. This assumes that the clinical context remained comparatively unchanged after April 1. The second part of the analysis thus tests whether the context entirely accounts for differences in incidence of psychiatric sequelae. Because HCOs take an average of 24 days to refresh their EHR on the TriNetX platform, we limited the follow-up period in these analyses to 60 days rather than 90 days to avoid relying on very few patients who had the index event shortly after April 1 when calculating the outcome probability at 90 days (since 114 days before August 1 correspond to April 9, 2020).

For the analysis of the psychiatric antecedents of COVID-19, the following three additional analyses were conducted:

Three-year window: The rationale for this analysis was to assess whether the increased incidence in COVID-19 diagnoses among patients with a history of psychiatric illness is specific to patients with a current or very recent psychiatric illness or appears to generalise to those with a longer history. In this analysis, the cohorts were defined in the exact same way as the cohorts for the primary analysis except that we allowed for a psychiatric diagnosis (ICD-10 codes F20-F48) to have been present anytime between January 21, 2017 and January 20, 2020. Similarly, the control cohort had to have had a health visit or any kind between January 21, 2017 and January 20, 2020. As this substantially increases the sample size and as the TriNetX platform allows for a maximum of 1.5 million individuals in a cohort to achieve propensity score matching, a finer stratification was used in this part of the analysis. A total of 26 strata were defined: 18–20 years old, 21–25, 26–30, 31–35, 36–40, 41–45, 46–50, 51–55, 56–60, 61–65, 66–70, 71–75, and 76 years old and older separately in males and females.

New diagnosis: The rationale for this additional analysis was to test whether differences in the incidence of COVID-19 diagnoses among people with a recent psychiatric illness is mostly the case for patients with a new diagnosis of psychiatric illness rather than any psychiatric illness made in the past year. To do so, we restricted the cohort of patients with a recent psychiatric illness to those who had a diagnosis (F20-F48) between January 21, 2019 and January 20, 2020 but who did not have any such diagnosis recorded before.

No physical comorbidity: The rationale for this additional analysis was to assess whether differences in the incidence of COVID-19 diagnoses among people with a recent psychiatric illness is associated with comorbidities that are known to be risk factors for COVID-19. Matching cohorts based on these comorbidities strongly limit their potential influence on the differential incidence of COVID-19 diagnoses. However this influence of comorbidities remains in two ways. First it might be that patients with a psychiatric illness have more severe comorbidities (matching for dichotomous variables does not control for this). Second it might be that comorbidities are distributed differently in the two cohorts even after matching (e.g. one cohort might have patients with comorbidities A and B and patients with comorbidity C, while the other have patients with comorbidities A and C and patients with comorbidity B). Furthermore, it might be that patients with comorbidities are contributing most to the differential incidence of COVID-19. To eliminate the influence of comorbidities on the findings, we excluded from the cohorts all those with comorbidities (Item 4-11 above). In terms of blood pressure, besides excluding those with a diagnosis of hypertension, we also excluded all those with a recorded systolic blood pressure above 140mmHg or a diastolic blood pressure above 90mmHg at any one time. In terms of BMI, besides excluding those with a diagnosis of overweight or obesity, we also excluded all those with a BMI recorded equal or above 25kg/m^2 at any one time. We then conducted the same analysis as the primary analysis using these restricted cohorts.

Propensity score matching

The log rank tests used in the Kaplan-Meier analyses of psychiatric sequelae and the χ^2-tests used in the analysis of psychiatric antecedents do not take the matching process into account (i.e. they are not paired). This is a limitation of the TriNetX platform. Monte Carlo studies have shown that this results in wider confidence intervals and lower Type I error rates than their nominal value[7]. The results are therefore likely more conservative than they would otherwise be if matching was accounted for. Some false negative findings might therefore remain in the paper. We consider this not to be a major issue since the large size of the sample combined with the relatively high threshold on the p-value imply that any false negative finding would likely have too small an effect size to be clinically significant.

Piecewise constant hazard ratio

The proportional hazard assumption was tested using the generalized Schoenfeld approach. Various alternative models have been proposed to deal with situations wherein the hazard ratio is not constant[8–10]. Here we used a piecewise constant model in which the hazard ratio is separately estimated for different time periods. The main advantage of this approach is that it is very informative[8] for it provides the actual value of the hazard ratio estimated for different time windows.

Specifically, when there was evidence of non-proportionality, we split the follow-up period in two phases: the 'early' phase (before 45 days) and the 'late' phase (from 45 days to 90 days). This was achieved by using a stepwise time function and replacing the 'cohort' indicator variable by the interaction between 'cohort' and 'phase' in the Cox model[11]. Separate hazard ratios (and corresponding p-values) were thereby estimated in the two phases using a proportional hazard model. The assumption that the hazard was proportional when accounting for the two phases was tested in exactly the same way as it was tested for the original model (i.e. using the generalized Schoenfeld approach implemented in the cox.zph function of the survival package in R).

Testing the trend of the association between relative risk and age

The relative risk for COVID-19 among patients with a recent diagnosis of psychiatric illness appeared to increase with age. This would imply that having a recent psychiatric diagnosis is a more important risk factor among older people. To test whether this observation can be explained by chance alone, we used a logistic regression.

In this regression, the dependent variable was the presence or absence of a COVID-19 diagnosis (so each of the 3,459,674 patients provided one value) and the independent variables were the presence or absence of a recent psychiatric diagnosis, sex, age (as an ordinal variable with the 5 categories used for stratification), and an interaction term between age and psychiatric diagnosis. The latter variable is the term of interest in the logistic regression. To capture the ordinal nature of the age variable, the variable was transformed with orthogonal polynomials (as is achieved in the glm function in R when the variable is labelled as ordinal). This results in values for linear, quadratic, cubic, and 4th power trends. We focused on the linear trend. We reported the adjusted odd ratios (OR) as the exponential of the logistic regression coefficient corresponding to the interaction term.

Supplementary Results

Table 1 – Characteristics of the COVID-19 and influenza cohorts before and after matching

	Before matching			After matching		
	COVID-19	**Influenza**	**SMD**	**COVID-19**	**Influenza**	**SMD**
Number	44779	43370	-	26497	26497	-
Age; mean (SD); y	47.1 (19.2)	35.3 (19.6)	0.609	42.7 (17.9)	41.8 (19.8)	0.0492
Sex						
Female	23477 (52.4)	22873 (52.7)	0.00622	14245 (53.8)	14311 (54)	0.005
Male	21071 (47.1)	20354 (46.9)	0.00249	12122 (45.7)	12065 (45.5)	0.00432
Other	231 (0.516)	143 (0.33)	0.0287	130 (0.491)	121 (0.457)	0.00495
Race						
White	20912 (46.7)	28061 (64.7)	0.368	14166 (53.5)	14756 (55.7)	0.0447
Black or African American	10885 (24.3)	8224 (19)	0.13	6617 (25)	6175 (23.3)	0.039
Asian	1284 (2.87)	1246 (2.87)	0.000332	847 (3.2)	848 (3.2)	0.000214
American Indian or Alaska Native	234 (0.523)	171 (0.394)	0.019	152 (0.574)	131 (0.494)	0.0109
Native Hawaiian or Other Pacific Islander	84 (0.188)	93 (0.214)	0.00599	60 (0.226)	55 (0.208)	0.00406
Unknown	11380 (25.4)	5575 (12.9)	0.323	4655 (17.6)	4532 (17.1)	0.0123
SBP						
At most 140 mm[Hg]	25643	36756	0.636	21096	20266	0.0757
140-160 mm[Hg]	12919	13667	0.058	9717	9266	0.0355
At least 160 mm[Hg]	6112	5563	0.0243	4289	4058	0.0239
DBP						
At most 90 mm[Hg]	26723	37612	0.641	21814	20952	0.0825
90-100 mm[Hg]	9574	10581	0.0718	7409	7037	0.0315
At least 100 mm[Hg]	4262	4408	0.0217	3163	3027	0.016
Hypertensive diseases (deprecated 2018)	12107 (27)	7828 (18)	0.216	7137 (26.9)	6444 (24.3)	0.0599
Hypertensive diseases	12124 (27.1)	7833 (18.1)	0.217	7140 (26.9)	6449 (24.3)	0.0597
Hypertensive chronic kidney disease	1342 (3)	821 (1.89)	0.0715	777 (2.93)	688 (2.6)	0.0205
BMI	15453 (34.5)	19217 (44.3)	-	12291 (46.4)	11765 (44.4)	-
At most 25 kg/m2	6304	10423	0.256	5334	5239	0.00897
25-30 kg/m2	7080	7885	0.0631	5502	5495	0.000651
At least 30 kg/m2	7703	7345	0.00709	5782	5519	0.0242
Overweight and obesity	6508 (14.5)	4986 (11.5)	0.0904	4185 (15.8)	3807 (14.4)	0.0399
Neoplasms	6638 (14.8)	5945 (13.7)	0.0319	4544 (17.1)	4327 (16.3)	0.0219
Haematological neoplasm	393 (0.878)	439 (1.01)	0.0139	298 (1.12)	302 (1.14)	0.00143
Certain disorders involving the immune mechanism	686 (1.53)	706 (1.63)	0.00769	486 (1.83)	490 (1.85)	0.00112
Type 1 diabetes mellitus	710 (1.59)	545 (1.26)	0.0278	436 (1.64)	404 (1.52)	0.00967
Type 2 diabetes mellitus	6335 (14.1)	3259 (7.51)	0.215	3297 (12.4)	2873 (10.8)	0.0499
Vascular dementia	163 (0.364)	45 (0.104)	0.0539	50 (0.189)	45 (0.17)	0.00446
Dementia in other diseases classified elsewhere	277 (0.619)	67 (0.154)	0.0748	77 (0.291)	65 (0.245)	0.00876
Unspecified dementia	665 (1.48)	144 (0.332)	0.122	165 (0.623)	142 (0.536)	0.0114
Alzheimer disease	240 (0.536)	51 (0.118)	0.0734	60 (0.226)	50 (0.189)	0.00829
Nicotine dependence	2248 (5.02)	3402 (7.84)	0.115	1893 (7.14)	1844 (6.96)	0.00722
Ischemic heart diseases	3144 (7.02)	1978 (4.56)	0.105	1814 (6.85)	1638 (6.18)	0.0269
Other forms of heart disease	6305 (14.1)	4656 (10.7)	0.102	3844 (14.5)	3546 (13.4)	0.0325
Cerebral infarction	824 (1.84)	433 (0.998)	0.0712	422 (1.59)	371 (1.4)	0.0159
Bronchitis; not specified as acute or chronic	1447 (3.23)	2706 (6.24)	0.142	1235 (4.66)	1242 (4.69)	0.00125

Simple and mucopurulent chronic bronchitis	116 (0.259)	157 (0.362)	0.0185	99 (0.374)	101 (0.381)	0.00123
Unspecified chronic bronchitis	117 (0.261)	146 (0.337)	0.0138	95 (0.359)	94 (0.355)	0.000633
Emphysema	496 (1.11)	398 (0.918)	0.019	319 (1.2)	303 (1.14)	0.00561
Other chronic obstructive pulmonary disease	1458 (3.26)	1397 (3.22)	0.00197	1022 (3.86)	942 (3.56)	0.016
Asthma	3568 (7.97)	5649 (13)	0.166	2876 (10.9)	2821 (10.6)	0.0067
Bronchiectasis	157 (0.351)	198 (0.457)	0.0167	114 (0.43)	114 (0.43)	0
Alcoholic liver disease	135 (0.301)	70 (0.161)	0.0292	73 (0.276)	58 (0.219)	0.0114
Hepatic failure; not elsewhere classified	196 (0.438)	121 (0.279)	0.0266	106 (0.4)	100 (0.377)	0.00364
Chronic hepatitis; not elsewhere classified	34 (0.076)	34 (0.078)	0.000888	26 (0.098)	23 (0.087)	0.00373
Fibrosis and cirrhosis of liver	358 (0.799)	225 (0.519)	0.0347	209 (0.789)	193 (0.728)	0.00696
Fatty (change of) liver; not elsewhere classified	1000 (2.23)	791 (1.82)	0.029	657 (2.48)	604 (2.28)	0.0131
Chronic passive congestion of liver	168 (0.375)	166 (0.383)	0.00123	128 (0.483)	116 (0.438)	0.00669
Portal hypertension	131 (0.293)	77 (0.178)	0.0238	75 (0.283)	65 (0.245)	0.00735
Other specified diseases of liver	621 (1.39)	588 (1.36)	0.00267	447 (1.69)	414 (1.56)	0.00985
Psoriasis	317 (0.708)	334 (0.77)	0.00726	223 (0.842)	211 (0.796)	0.00503
Rheumatoid arthritis with rheumatoid factor	154 (0.344)	135 (0.311)	0.00571	102 (0.385)	98 (0.37)	0.00246
Other rheumatoid arthritis	439 (0.98)	382 (0.881)	0.0104	277 (1.04)	279 (1.05)	0.000741
Systemic lupus erythematosus (SLE)	202 (0.451)	182 (0.42)	0.00478	143 (0.54)	134 (0.506)	0.00471
Chronic kidney disease (CKD)	2656 (5.93)	1528 (3.52)	0.114	1452 (5.48)	1284 (4.85)	0.0287
Renal Transplantation Procedures	65 (0.145)	65 (0.15)	0.00123	48 (0.181)	45 (0.17)	0.00271
Liver Transplantation Procedures	19 (0.042)	12 (0.028)	0.00789	10 (0.038)	10 (0.038)	0

SMD=Standardized Mean Difference, SBP=Systolic Blood Pressure, DBP=Diastolic Blood Pressure, BMI=Body Mass Index

Table 2 – Characteristics of the COVID-19 and other respiratory tract infection (RTI) cohorts before and after matching

	Before matching			After matching		
	COVID-19	**Other RTI**	**SMD**	**COVID-19**	**Other RTI**	**SMD**
Number	44779	324281	-	44775	44775	-
Age; mean (SD); y	47.1 (19.2)	40.5 (21.1)	0.325	47.1 (19.2)	47.2 (19.9)	0.00267
Sex						
Female	23477 (52.4)	182154 (56.2)	0.0752	23475 (52.4)	23676 (52.9)	0.00899
Male	21071 (47.1)	140376 (43.3)	0.0758	21069 (47.1)	20865 (46.6)	0.00913
Other	231 (0.516)	1751 (0.54)	0.00333	231 (0.516)	234 (0.523)	0.000932
Race						
White	20912 (46.7)	212872 (65.6)	0.389	20912 (46.7)	21924 (49)	0.0453
Black or African American	10885 (24.3)	57161 (17.6)	0.165	10883 (24.3)	10081 (22.5)	0.0423
Asian	1284 (2.87)	8114 (2.5)	0.0226	1284 (2.87)	1276 (2.85)	0.00107
American Indian or Alaska Native	234 (0.523)	1303 (0.402)	0.0178	234 (0.523)	224 (0.5)	0.00313
Native Hawaiian or Other Pacific Islander	84 (0.188)	576 (0.178)	0.00233	84 (0.188)	62 (0.138)	0.0122
Unknown	11380 (25.4)	44255 (13.6)	0.3	11378 (25.4)	11208 (25)	0.00874
SBP						
At most 140 mm[Hg]	25643	254653	0.468	25643	24698	0.0426
140-160 mm[Hg]	12919	107854	0.0954	12919	12541	0.0187
At least 160 mm[Hg]	6112	47847	0.0317	6112	5744	0.0243
DBP						
At most 90 mm[Hg]	26723	261260	0.469	26723	25742	0.0445
90-100 mm[Hg]	9574	82431	0.0955	9574	9299	0.0151
At least 100 mm[Hg]	4262	35031	0.0425	4262	4001	0.0201
Hypertensive diseases (deprecated 2018)	12107 (27)	76545 (23.6)	0.079	12106 (27)	11329 (25.3)	0.0395
Hypertensive diseases	12124 (27.1)	76617 (23.6)	0.0793	12123 (27.1)	11340 (25.3)	0.0398
Hypertensive chronic kidney disease	1342 (3)	7342 (2.26)	0.0458	1342 (3)	1159 (2.59)	0.0248
BMI						
At most 25 kg/m2	6304	71826	0.211	6304	6092	0.0137
25-30 kg/m2	7080	62408	0.0904	7080	6941	0.00854
At least 30 kg/m2	7703	59322	0.0286	7703	7534	0.01
Overweight and obesity	6508 (14.5)	43143 (13.3)	0.0355	6508 (14.5)	6082 (13.6)	0.0274
Neoplasms	6638 (14.8)	56723 (17.5)	0.0725	6638 (14.8)	6617 (14.8)	0.00132
Haematological neoplasm	393 (0.878)	3858 (1.19)	0.0309	393 (0.878)	395 (0.882)	0.000478
Certain disorders involving the immune mechanism	686 (1.53)	5691 (1.76)	0.0175	686 (1.53)	641 (1.43)	0.00832
Type 1 diabetes mellitus	710 (1.59)	4193 (1.29)	0.0246	710 (1.59)	648 (1.45)	0.0113
Type 2 diabetes mellitus	6335 (14.1)	30602 (9.44)	0.146	6334 (14.1)	5625 (12.6)	0.0466
Vascular dementia	163 (0.364)	484 (0.149)	0.0425	163 (0.364)	126 (0.281)	0.0146
Dementia in other diseases classified elsewhere	277 (0.619)	749 (0.231)	0.0596	276 (0.616)	180 (0.402)	0.0301
Unspecified dementia	665 (1.48)	1798 (0.554)	0.0927	664 (1.48)	497 (1.11)	0.033
Alzheimer disease	240 (0.536)	660 (0.204)	0.0548	239 (0.534)	165 (0.369)	0.0247
Nicotine dependence	2248 (5.02)	27746 (8.56)	0.141	2248 (5.02)	2314 (5.17)	0.0067
Ischemic heart diseases	3144 (7.02)	20928 (6.45)	0.0226	3144 (7.02)	2823 (6.3)	0.0288
Other forms of heart disease	6305 (14.1)	44180 (13.6)	0.0132	6305 (14.1)	5963 (13.3)	0.0222
Cerebral infarction	824 (1.84)	4450 (1.37)	0.0372	824 (1.84)	730 (1.63)	0.0161
Bronchitis; not specified as acute or chronic	1447 (3.23)	23067 (7.11)	0.176	1447 (3.23)	1447 (3.23)	0
Simple and mucopurulent chronic bronchitis	116 (0.259)	1624 (0.501)	0.0393	116 (0.259)	135 (0.302)	0.00803
Unspecified chronic bronchitis	117 (0.261)	1725 (0.532)	0.0431	117 (0.261)	127 (0.284)	0.00428
Emphysema	496 (1.11)	5003 (1.54)	0.0381	496 (1.11)	481 (1.07)	0.00322

Other chronic obstructive pulmonary disease	1458 (3.26)	14526 (4.48)	0.0635	1458 (3.26)	1356 (3.03)	0.0131
Asthma	3568 (7.97)	41086 (12.7)	0.155	3568 (7.97)	3372 (7.53)	0.0164
Bronchiectasis	157 (0.351)	2401 (0.74)	0.0529	157 (0.351)	151 (0.337)	0.00229
Alcoholic liver disease	135 (0.301)	948 (0.292)	0.00168	135 (0.302)	126 (0.281)	0.00373
Hepatic failure; not elsewhere classified	196 (0.438)	1188 (0.366)	0.0113	196 (0.438)	153 (0.342)	0.0154
Chronic hepatitis; not elsewhere classified	34 (0.076)	254 (0.078)	0.000864	34 (0.076)	18 (0.04)	0.0148
Fibrosis and cirrhosis of liver	358 (0.799)	2248 (0.693)	0.0123	358 (0.8)	283 (0.632)	0.0199
Fatty (change of) liver; not elsewhere classified	1000 (2.23)	7071 (2.18)	0.00359	1000 (2.23)	896 (2)	0.0161
Chronic passive congestion of liver	168 (0.375)	1411 (0.435)	0.00944	168 (0.375)	159 (0.355)	0.00333
Portal hypertension	131 (0.293)	1012 (0.312)	0.00356	131 (0.293)	116 (0.259)	0.00639
Other specified diseases of liver	621 (1.39)	5070 (1.56)	0.0147	621 (1.39)	558 (1.25)	0.0123
Psoriasis	317 (0.708)	3293 (1.01)	0.0333	317 (0.708)	314 (0.701)	0.000801
Rheumatoid arthritis with rheumatoid factor	154 (0.344)	1343 (0.414)	0.0114	154 (0.344)	154 (0.344)	0
Other rheumatoid arthritis	439 (0.98)	3759 (1.16)	0.0174	439 (0.98)	403 (0.9)	0.00833
Systemic lupus erythematosus (SLE)	202 (0.451)	1408 (0.434)	0.00255	202 (0.451)	186 (0.415)	0.00544
Chronic kidney disease (CKD)	2656 (5.93)	15156 (4.67)	0.0561	2656 (5.93)	2329 (5.2)	0.0319
Renal Transplantation Procedures	65 (0.145)	369 (0.114)	0.00872	65 (0.145)	54 (0.121)	0.00674
Liver Transplantation Procedures	19 (0.042)	71 (0.022)	0.0115	19 (0.042)	11 (0.025)	0.00976

SMD=Standardized Mean Difference, SBP=Systolic Blood Pressure, DBP=Diastolic Blood Pressure, BMI=Body Mass Index

Table 3 – Characteristics of the COVID-19 and skin infection cohorts before and after matching

	Before matching			After matching		
	COVID-19	**Skin infection**	**SMD**	**COVID-19**	**Skin infection**	**SMD**
Number	44779	75076	-	38977	38977	-
Age; mean (SD); y	47.1 (19.2)	47.9 (21.1)	0.0413	46.8 (19.1)	46.7 (20.9)	0.00501
Sex						
Female	23477 (52.4)	33694 (44.9)	0.151	19855 (50.9)	19956 (51.2)	0.00518
Male	21071 (47.1)	40854 (54.4)	0.148	18894 (48.5)	18791 (48.2)	0.00529
Other	231 (0.516)	528 (0.703)	0.0241	228 (0.585)	230 (0.59)	0.000671
Race						
White	20912 (46.7)	51454 (68.5)	0.453	20645 (53)	20775 (53.3)	0.00668
Black or African American	10885 (24.3)	13139 (17.5)	0.168	9346 (24)	9296 (23.9)	0.00301
Asian	1284 (2.87)	1244 (1.66)	0.0815	1024 (2.63)	1033 (2.65)	0.00144
American Indian or Alaska Native	234 (0.523)	358 (0.477)	0.00648	218 (0.559)	220 (0.564)	0.000686
Native Hawaiian or Other Pacific Islander	84 (0.188)	102 (0.136)	0.0129	76 (0.195)	68 (0.174)	0.00478
Unknown	11380 (25.4)	8779 (11.7)	0.359	7668 (19.7)	7585 (19.5)	0.00537
SBP						
At most 140 mm[Hg]	25643	56254	0.38	24708	24609	0.00527
140-160 mm[Hg]	12919	33533	0.333	12602	12393	0.0115
At least 160 mm[Hg]	6112	18100	0.27	6000	5844	0.0112
DBP						
At most 90 mm[Hg]	26723	58709	0.409	25806	25675	0.0071
90-100 mm[Hg]	9574	24808	0.264	9272	9190	0.00495
At least 100 mm[Hg]	4262	11944	0.193	4138	4053	0.00711
Hypertensive diseases (deprecated 2018)	12107 (27)	27013 (36)	0.193	11306 (29)	11177 (28.7)	0.00731
Hypertensive diseases	12124 (27.1)	27038 (36)	0.193	11322 (29)	11190 (28.7)	0.00747
Hypertensive chronic kidney disease	1342 (3)	3774 (5.03)	0.104	1285 (3.3)	1274 (3.27)	0.00158
BMI						
At most 25 kg/m2	6304	14323	0.135	6119	6368	0.0174
25-30 kg/m2	7080	15471	0.125	6740	6887	0.00993
At least 30 kg/m2	7703	17710	0.159	7358	7317	0.00269
Overweight and obesity	6508 (14.5)	14408 (19.2)	0.125	6151 (15.8)	6136 (15.7)	0.00106
Neoplasms	6638 (14.8)	17628 (23.5)	0.221	6443 (16.5)	6583 (16.9)	0.00963
Haematological neoplasm	393 (0.878)	1265 (1.68)	0.0718	386 (0.99)	407 (1.04)	0.00537
Certain disorders involving the immune mechanism	686 (1.53)	1695 (2.26)	0.0532	662 (1.7)	643 (1.65)	0.0038
Type 1 diabetes mellitus	710 (1.59)	2594 (3.46)	0.119	700 (1.8)	707 (1.81)	0.00135
Type 2 diabetes mellitus	6335 (14.1)	15013 (20)	0.156	5967 (15.3)	5788 (14.8)	0.0128
Vascular dementia	163 (0.364)	161 (0.214)	0.0279	123 (0.316)	123 (0.316)	0
Dementia in other diseases classified elsewhere	277 (0.619)	319 (0.425)	0.0269	227 (0.582)	219 (0.562)	0.00272
Unspecified dementia	665 (1.48)	699 (0.931)	0.0507	508 (1.3)	500 (1.28)	0.00182
Alzheimer disease	240 (0.536)	262 (0.349)	0.0282	184 (0.472)	189 (0.485)	0.00186
Nicotine dependence	2248 (5.02)	10500 (14)	0.309	2244 (5.76)	2352 (6.03)	0.0118
Ischemic heart diseases	3144 (7.02)	8122 (10.8)	0.134	2973 (7.63)	2928 (7.51)	0.00436
Other forms of heart disease	6305 (14.1)	15399 (20.5)	0.171	5979 (15.3)	5893 (15.1)	0.00614
Cerebral infarction	824 (1.84)	1672 (2.23)	0.0274	742 (1.9)	736 (1.89)	0.00113
Bronchitis; not specified as acute or chronic	1447 (3.23)	3567 (4.75)	0.0777	1385 (3.55)	1321 (3.39)	0.00897
Simple and mucopurulent chronic bronchitis	116 (0.259)	406 (0.541)	0.0447	113 (0.29)	110 (0.282)	0.00144
Unspecified chronic bronchitis	117 (0.261)	356 (0.474)	0.0352	111 (0.285)	107 (0.275)	0.00194
Emphysema	496 (1.11)	1324 (1.76)	0.0552	473 (1.21)	482 (1.24)	0.0021
Other chronic obstructive pulmonary disease	1458 (3.26)	3996 (5.32)	0.102	1371 (3.52)	1321 (3.39)	0.00703

Asthma	3568 (7.97)	6972 (9.29)	0.047	3311 (8.49)	3272 (8.4)	0.0036
Bronchiectasis	157 (0.351)	338 (0.45)	0.0158	150 (0.385)	140 (0.359)	0.00421
Alcoholic liver disease	135 (0.301)	553 (0.737)	0.0606	133 (0.341)	137 (0.351)	0.00175
Hepatic failure; not elsewhere classified	196 (0.438)	527 (0.702)	0.0351	180 (0.462)	177 (0.454)	0.00114
Chronic hepatitis; not elsewhere classified	34 (0.076)	117 (0.156)	0.0235	33 (0.085)	27 (0.069)	0.00555
Fibrosis and cirrhosis of liver	358 (0.799)	1078 (1.44)	0.0606	348 (0.893)	333 (0.854)	0.00414
Fatty (change of) liver; not elsewhere classified	1000 (2.23)	2307 (3.07)	0.0523	953 (2.44)	899 (2.31)	0.0091
Chronic passive congestion of liver	168 (0.375)	457 (0.609)	0.0334	166 (0.426)	155 (0.398)	0.00441
Portal hypertension	131 (0.293)	481 (0.641)	0.0511	129 (0.331)	121 (0.31)	0.00363
Other specified diseases of liver	621 (1.39)	1673 (2.23)	0.0632	600 (1.54)	615 (1.58)	0.00311
Psoriasis	317 (0.708)	1352 (1.8)	0.0983	314 (0.806)	309 (0.793)	0.00144
Rheumatoid arthritis with rheumatoid factor	154 (0.344)	463 (0.617)	0.0395	147 (0.377)	141 (0.362)	0.00254
Other rheumatoid arthritis	439 (0.98)	1298 (1.73)	0.0648	420 (1.08)	400 (1.03)	0.00503
Systemic lupus erythematosus (SLE)	202 (0.451)	452 (0.602)	0.0209	194 (0.498)	190 (0.487)	0.00147
Chronic kidney disease (CKD)	2656 (5.93)	7019 (9.35)	0.129	2536 (6.51)	2485 (6.38)	0.00533
Renal Transplantation Procedures	65 (0.145)	180 (0.24)	0.0216	65 (0.167)	64 (0.164)	0.000631
Liver Transplantation Procedures	19 (0.042)	53 (0.071)	0.0119	18 (0.046)	17 (0.044)	0.00121

SMD=Standardized Mean Difference, SBP=Systolic Blood Pressure, DBP=Diastolic Blood Pressure, BMI=Body Mass Index

Table 4 – Characteristics of the COVID-19 and cholelithiasis cohorts before and after matching

	Before matching			After matching		
	COVID-19	**Cholelithiasis**	**SMD**	**COVID-19**	**Cholelithiasis**	**SMD**
Number	44779	26102	-	19733	19733	-
Age; mean (SD); y	47.1 (19.2)	56 (18.9)	0.468	53.6 (19)	53.2 (19.3)	0.024
Sex						
Female	23477 (52.4)	14947 (57.3)	0.0973	12241 (62)	11637 (59)	0.0626
Male	21071 (47.1)	10620 (40.7)	0.129	7266 (36.8)	7862 (39.8)	0.0622
Other	231 (0.516)	535 (2.05)	0.137	226 (1.14)	234 (1.19)	0.00378
Race						
White	20912 (46.7)	16991 (65.1)	0.377	12041 (61)	12033 (61)	0.000831
Black or African American	10885 (24.3)	4011 (15.4)	0.226	3533 (17.9)	3472 (17.6)	0.00809
Asian	1284 (2.87)	708 (2.71)	0.00941	539 (2.73)	550 (2.79)	0.0034
American Indian or Alaska Native	234 (0.523)	88 (0.337)	0.0283	79 (0.4)	75 (0.38)	0.00325
Native Hawaiian or Other Pacific Islander	84 (0.188)	47 (0.18)	0.00176	36 (0.182)	35 (0.177)	0.0012
Unknown	11380 (25.4)	4257 (16.3)	0.225	3505 (17.8)	3568 (18.1)	0.00832
SBP						
At most 140 mm[Hg]	25643	17992	0.244	12995	12687	0.0327
140-160 mm[Hg]	12919	12411	0.392	7995	7835	0.0165
At least 160 mm[Hg]	6112	7443	0.371	4366	4264	0.0125
DBP						
At most 90 mm[Hg]	26723	18850	0.267	13678	13331	0.0378
90-100 mm[Hg]	9574	8903	0.287	5793	5685	0.0121
At least 100 mm[Hg]	4262	4483	0.227	2729	2725	0.000587
Hypertensive diseases (deprecated 2018)	12107 (27)	11855 (45.4)	0.39	7699 (39)	7548 (38.3)	0.0157
Hypertensive diseases	12124 (27.1)	11873 (45.5)	0.39	7711 (39.1)	7558 (38.3)	0.0159
Hypertensive chronic kidney disease	1342 (3)	1863 (7.14)	0.19	972 (4.93)	977 (4.95)	0.00117
BMI						
At most 25 kg/m2	6304	4633	0.1	3049	3092	0.00601
25-30 kg/m2	7080	6871	0.26	4302	4226	0.00936
At least 30 kg/m2	7703	7979	0.317	5159	4938	0.0257
Overweight and obesity	6508 (14.5)	5778 (22.1)	0.197	3929 (19.9)	3810 (19.3)	0.0152
Neoplasms	6638 (14.8)	8533 (32.7)	0.429	4962 (25.1)	4826 (24.5)	0.016
Haematological neoplasm	393 (0.878)	620 (2.38)	0.119	330 (1.67)	332 (1.68)	0.000789
Certain disorders involving the immune mechanism	686 (1.53)	756 (2.9)	0.0928	413 (2.09)	415 (2.1)	0.000707
Type 1 diabetes mellitus	710 (1.59)	621 (2.38)	0.057	393 (1.99)	413 (2.09)	0.00717
Type 2 diabetes mellitus	6335 (14.1)	5517 (21.1)	0.184	3612 (18.3)	3626 (18.4)	0.00183
Vascular dementia	163 (0.364)	75 (0.287)	0.0135	71 (0.36)	62 (0.314)	0.00787
Dementia in other diseases classified elsewhere	277 (0.619)	161 (0.617)	0.000228	136 (0.689)	137 (0.694)	0.000611
Unspecified dementia	665 (1.48)	322 (1.23)	0.0217	318 (1.61)	284 (1.44)	0.0141
Alzheimer disease	240 (0.536)	148 (0.567)	0.00419	120 (0.608)	120 (0.608)	0
Nicotine dependence	2248 (5.02)	2923 (11.2)	0.228	1648 (8.35)	1643 (8.33)	0.000916
Ischemic heart diseases	3144 (7.02)	4472 (17.1)	0.314	2476 (12.5)	2458 (12.5)	0.00276
Other forms of heart disease	6305 (14.1)	7523 (28.8)	0.365	4496 (22.8)	4411 (22.4)	0.0103
Cerebral infarction	824 (1.84)	855 (3.28)	0.091	553 (2.8)	543 (2.75)	0.00308
Bronchitis; not specified as acute or chronic	1447 (3.23)	974 (3.73)	0.0273	659 (3.34)	683 (3.46)	0.00671
Simple and mucopurulent chronic bronchitis	116 (0.259)	154 (0.59)	0.0509	79 (0.4)	77 (0.39)	0.00162
Unspecified chronic bronchitis	117 (0.261)	145 (0.556)	0.0461	83 (0.421)	67 (0.34)	0.0132
Emphysema	496 (1.11)	828 (3.17)	0.143	389 (1.97)	409 (2.07)	0.0072
Other chronic obstructive pulmonary disease	1458 (3.26)	1759 (6.74)	0.16	1048 (5.31)	1025 (5.19)	0.00522

Asthma	3568 (7.97)	2189 (8.39)	0.0153	1704 (8.64)	1641 (8.32)	0.0115
Bronchiectasis	157 (0.351)	332 (1.27)	0.103	133 (0.674)	138 (0.699)	0.00307
Alcoholic liver disease	135 (0.301)	917 (3.51)	0.236	128 (0.649)	174 (0.882)	0.0268
Hepatic failure; not elsewhere classified	196 (0.438)	991 (3.8)	0.235	176 (0.892)	241 (1.22)	0.0322
Chronic hepatitis; not elsewhere classified	34 (0.076)	143 (0.548)	0.0847	29 (0.147)	35 (0.177)	0.00756
Fibrosis and cirrhosis of liver	358 (0.799)	2063 (7.9)	0.354	345 (1.75)	457 (2.32)	0.0402
Fatty (change of) liver; not elsewhere classified	1000 (2.23)	3274 (12.5)	0.402	982 (4.98)	1189 (6.02)	0.046
Chronic passive congestion of liver	168 (0.375)	528 (2.02)	0.152	153 (0.775)	167 (0.846)	0.00791
Portal hypertension	131 (0.293)	1283 (4.92)	0.293	130 (0.659)	224 (1.14)	0.0505
Other specified diseases of liver	621 (1.39)	2237 (8.57)	0.335	594 (3.01)	763 (3.87)	0.047
Psoriasis	317 (0.708)	358 (1.37)	0.0655	220 (1.12)	201 (1.02)	0.00937
Rheumatoid arthritis with rheumatoid factor	154 (0.344)	136 (0.521)	0.027	87 (0.441)	99 (0.502)	0.00888
Other rheumatoid arthritis	439 (0.98)	507 (1.94)	0.0802	325 (1.65)	322 (1.63)	0.0012
Systemic lupus erythematosus (SLE)	202 (0.451)	149 (0.571)	0.0168	104 (0.527)	100 (0.507)	0.00283
Chronic kidney disease (CKD)	2656 (5.93)	3325 (12.7)	0.236	1890 (9.58)	1861 (9.43)	0.00501
Renal Transplantation Procedures	65 (0.145)	62 (0.238)	0.0211	39 (0.198)	38 (0.193)	0.00115
Liver Transplantation Procedures	19 (0.042)	51 (0.195)	0.0444	17 (0.086)	19 (0.096)	0.00336

SMD=Standardized Mean Difference, SBP=Systolic Blood Pressure, DBP=Diastolic Blood Pressure, BMI=Body Mass Index

Table 5 – Characteristics of the COVID-19 and urolithiasis cohorts before and after matching

	Before matching			After matching		
	COVID-19	**Urolithiasis**	**SMD**	**COVID-19**	**Urolithiasis**	**SMD**
Number	44779	50964	-	28827	28827	-
Age; mean (SD); y	47.1 (19.2)	55.6 (17.6)	0.459	50.3 (19.1)	50.2 (18)	0.00921
Sex						
Female	23477 (52.4)	21005 (41.2)	0.226	13966 (48.4)	14034 (48.7)	0.00472
Male	21071 (47.1)	29345 (57.6)	0.212	14641 (50.8)	14569 (50.5)	0.005
Other	231 (0.516)	614 (1.2)	0.0746	220 (0.763)	224 (0.777)	0.00159
Race						
White	20912 (46.7)	38930 (76.4)	0.641	18353 (63.7)	18431 (63.9)	0.00563
Black or African American	10885 (24.3)	4625 (9.07)	0.417	4382 (15.2)	4282 (14.9)	0.00971
Asian	1284 (2.87)	1161 (2.28)	0.0372	817 (2.83)	839 (2.91)	0.00457
American Indian or Alaska Native	234 (0.523)	141 (0.277)	0.039	141 (0.489)	126 (0.437)	0.00766
Native Hawaiian or Other Pacific Islander	84 (0.188)	66 (0.13)	0.0146	55 (0.191)	48 (0.167)	0.00575
Unknown	11380 (25.4)	6041 (11.9)	0.354	5079 (17.6)	5101 (17.7)	0.002
SBP						
At most 140 mm[Hg]	25643	38441	0.392	18714	18566	0.0107
140-160 mm[Hg]	12919	26944	0.504	10358	10165	0.014
At least 160 mm[Hg]	6112	14774	0.381	5042	4977	0.00595
DBP						
At most 90 mm[Hg]	26723	40100	0.421	19530	19381	0.011
90-100 mm[Hg]	9574	20298	0.409	7674	7586	0.00692
At least 100 mm[Hg]	4262	9765	0.278	3472	3406	0.00706
Hypertensive diseases (deprecated 2018)	12107 (27)	22179 (43.5)	0.35	9134 (31.7)	9141 (31.7)	0.000522
Hypertensive diseases	12124 (27.1)	22189 (43.5)	0.35	9145 (31.7)	9147 (31.7)	0.000149
Hypertensive chronic kidney disease	1342 (3)	3446 (6.76)	0.175	1130 (3.92)	1151 (3.99)	0.00374
BMI						
At most 25 kg/m2	6304	10734	0.184	5041	5038	0.000274
25-30 kg/m2	7080	15132	0.336	5802	5824	0.0019
At least 30 kg/m2	7703	15352	0.308	6060	5957	0.0088
Overweight and obesity	6508 (14.5)	9729 (19.1)	0.122	4409 (15.3)	4372 (15.2)	0.00357
Neoplasms	6638 (14.8)	17094 (33.5)	0.448	5755 (20)	5744 (19.9)	0.000955
Haematological neoplasm	393 (0.878)	938 (1.84)	0.0832	350 (1.21)	344 (1.19)	0.00191
Certain disorders involving the immune mechanism	686 (1.53)	1283 (2.52)	0.07	557 (1.93)	600 (2.08)	0.0106
Type 1 diabetes mellitus	710 (1.59)	1078 (2.12)	0.0393	499 (1.73)	511 (1.77)	0.00317
Type 2 diabetes mellitus	6335 (14.1)	9446 (18.5)	0.119	4382 (15.2)	4309 (14.9)	0.00708
Vascular dementia	163 (0.364)	89 (0.175)	0.0365	85 (0.295)	74 (0.257)	0.00728
Dementia in other diseases classified elsewhere	277 (0.619)	184 (0.361)	0.0369	156 (0.541)	134 (0.465)	0.0108
Unspecified dementia	665 (1.48)	343 (0.673)	0.0787	343 (1.19)	300 (1.04)	0.0142
Alzheimer disease	240 (0.536)	157 (0.308)	0.0352	133 (0.461)	110 (0.382)	0.0123
Nicotine dependence	2248 (5.02)	5001 (9.81)	0.184	1900 (6.59)	1979 (6.86)	0.0109
Ischemic heart diseases	3144 (7.02)	7092 (13.9)	0.227	2572 (8.92)	2545 (8.83)	0.00329
Other forms of heart disease	6305 (14.1)	11765 (23.1)	0.233	4812 (16.7)	4780 (16.6)	0.00298
Cerebral infarction	824 (1.84)	1394 (2.74)	0.0599	611 (2.12)	598 (2.07)	0.00315
Bronchitis; not specified as acute or chronic	1447 (3.23)	2188 (4.29)	0.0558	1016 (3.52)	1019 (3.54)	0.000564
Simple and mucopurulent chronic bronchitis	116 (0.259)	223 (0.438)	0.0303	84 (0.291)	90 (0.312)	0.00379
Unspecified chronic bronchitis	117 (0.261)	216 (0.424)	0.0278	81 (0.281)	86 (0.298)	0.00323
Emphysema	496 (1.11)	1010 (1.98)	0.0709	401 (1.39)	415 (1.44)	0.00411
Other chronic obstructive pulmonary disease	1458 (3.26)	2273 (4.46)	0.0625	1093 (3.79)	1102 (3.82)	0.00163

Asthma	3568 (7.97)	4206 (8.25)	0.0104	2306 (8)	2255 (7.82)	0.00655
Bronchiectasis	157 (0.351)	398 (0.781)	0.0574	132 (0.458)	131 (0.454)	0.000515
Alcoholic liver disease	135 (0.301)	264 (0.518)	0.0339	109 (0.378)	113 (0.392)	0.00224
Hepatic failure; not elsewhere classified	196 (0.438)	374 (0.734)	0.0388	151 (0.524)	151 (0.524)	0
Chronic hepatitis; not elsewhere classified	34 (0.076)	86 (0.169)	0.0266	26 (0.09)	33 (0.114)	0.00759
Fibrosis and cirrhosis of liver	358 (0.799)	843 (1.65)	0.0777	311 (1.08)	317 (1.1)	0.00201
Fatty (change of) liver; not elsewhere classified	1000 (2.23)	3619 (7.1)	0.232	941 (3.26)	951 (3.3)	0.00195
Chronic passive congestion of liver	168 (0.375)	647 (1.27)	0.0992	150 (0.52)	144 (0.5)	0.00292
Portal hypertension	131 (0.293)	420 (0.824)	0.0714	120 (0.416)	123 (0.427)	0.00161
Other specified diseases of liver	621 (1.39)	2561 (5.03)	0.208	582 (2.02)	605 (2.1)	0.00562
Psoriasis	317 (0.708)	812 (1.59)	0.0831	272 (0.944)	283 (0.982)	0.00391
Rheumatoid arthritis with rheumatoid factor	154 (0.344)	292 (0.573)	0.0339	112 (0.389)	116 (0.402)	0.00221
Other rheumatoid arthritis	439 (0.98)	996 (1.95)	0.0811	360 (1.25)	380 (1.32)	0.00616
Systemic lupus erythematosus (SLE)	202 (0.451)	278 (0.545)	0.0134	155 (0.538)	148 (0.513)	0.00336
Chronic kidney disease (CKD)	2656 (5.93)	6537 (12.8)	0.238	2227 (7.72)	2266 (7.86)	0.00505
Renal Transplantation Procedures	65 (0.145)	126 (0.247)	0.0231	61 (0.212)	59 (0.205)	0.00152
Liver Transplantation Procedures	19 (0.042)	45 (0.088)	0.0179	18 (0.062)	20 (0.069)	0.0027

SMD=Standardized Mean Difference, SBP=Systolic Blood Pressure, DBP=Diastolic Blood Pressure, BMI=Body Mass Index

Table 6 – Characteristics of the COVID-19 and fracture cohorts before and after matching

	Before matching			After matching		
	COVID-19	**Fracture**	**SMD**	**COVID-19**	**Fracture**	**SMD**
Number	44779	68547	-	37841	37841	-
Age; mean (SD); y	47.1 (19.2)	49.5 (24.7)	0.106	47.3 (19.5)	48 (23.6)	0.0291
Sex						
Female	23477 (52.4)	34263 (50)	0.0489	19251 (50.9)	19509 (51.6)	0.0136
Male	21071 (47.1)	33664 (49.1)	0.0411	18359 (48.5)	18083 (47.8)	0.0146
Other	231 (0.516)	620 (0.904)	0.0463	231 (0.61)	249 (0.658)	0.00599
Race						
White	20912 (46.7)	49742 (72.6)	0.547	20845 (55.1)	21049 (55.6)	0.0108
Black or African American	10885 (24.3)	8250 (12)	0.322	7401 (19.6)	7391 (19.5)	0.000666
Asian	1284 (2.87)	946 (1.38)	0.103	871 (2.3)	880 (2.33)	0.00158
American Indian or Alaska Native	234 (0.523)	236 (0.344)	0.0271	203 (0.536)	204 (0.539)	0.000361
Native Hawaiian or Other Pacific Islander	84 (0.188)	66 (0.096)	0.0243	65 (0.172)	63 (0.166)	0.00129
Unknown	11380 (25.4)	9307 (13.6)	0.302	8456 (22.3)	8254 (21.8)	0.0129
SBP						
At most 140 mm[Hg]	25643	45742	0.196	22371	21781	0.0316
140-160 mm[Hg]	12919	26924	0.221	11625	11352	0.0157
At least 160 mm[Hg]	6112	14874	0.212	5655	5406	0.0186
DBP						
At most 90 mm[Hg]	26723	48059	0.22	23409	22786	0.0338
90-100 mm[Hg]	9574	19031	0.149	8456	8228	0.0145
At least 100 mm[Hg]	4262	9327	0.128	3852	3716	0.012
Hypertensive diseases (deprecated 2018)	12107 (27)	20114 (29.3)	0.0513	10007 (26.4)	9895 (26.1)	0.00672
Hypertensive diseases	12124 (27.1)	20132 (29.4)	0.051	10023 (26.5)	9905 (26.2)	0.00708
Hypertensive chronic kidney disease	1342 (3)	2471 (3.6)	0.034	1157 (3.06)	1131 (2.99)	0.00401
BMI						
At most 25 kg/m2	6304	16713	0.264	6174	6187	0.000929
25-30 kg/m2	7080	14419	0.135	6475	6531	0.00392
At least 30 kg/m2	7703	12332	0.0207	6592	6638	0.0032
Overweight and obesity	6508 (14.5)	7470 (10.9)	0.109	4809 (12.7)	4842 (12.8)	0.00261
Neoplasms	6638 (14.8)	12492 (18.2)	0.0916	5801 (15.3)	5778 (15.3)	0.00169
Haematological neoplasm	393 (0.878)	888 (1.3)	0.0403	361 (0.954)	339 (0.896)	0.00607
Certain disorders involving the immune mechanism	686 (1.53)	882 (1.29)	0.0208	533 (1.41)	523 (1.38)	0.00225
Type 1 diabetes mellitus	710 (1.59)	1187 (1.73)	0.0114	603 (1.59)	612 (1.62)	0.00189
Type 2 diabetes mellitus	6335 (14.1)	7579 (11.1)	0.0932	4715 (12.5)	4688 (12.4)	0.00216
Vascular dementia	163 (0.364)	290 (0.423)	0.00943	143 (0.378)	147 (0.388)	0.00171
Dementia in other diseases classified elsewhere	277 (0.619)	619 (0.903)	0.0327	257 (0.679)	255 (0.674)	0.000645
Unspecified dementia	665 (1.48)	1414 (2.06)	0.0438	611 (1.62)	642 (1.7)	0.00642
Alzheimer disease	240 (0.536)	544 (0.794)	0.0317	223 (0.589)	217 (0.573)	0.00209
Nicotine dependence	2248 (5.02)	7382 (10.8)	0.214	2218 (5.86)	2330 (6.16)	0.0125
Ischemic heart diseases	3144 (7.02)	6232 (9.09)	0.0761	2782 (7.35)	2759 (7.29)	0.00233
Other forms of heart disease	6305 (14.1)	12492 (18.2)	0.113	5587 (14.8)	5499 (14.5)	0.00658
Cerebral infarction	824 (1.84)	1596 (2.33)	0.0342	728 (1.92)	753 (1.99)	0.00477
Bronchitis; not specified as acute or chronic	1447 (3.23)	1885 (2.75)	0.0283	1100 (2.91)	1112 (2.94)	0.00188
Simple and mucopurulent chronic bronchitis	116 (0.259)	256 (0.373)	0.0204	101 (0.267)	92 (0.243)	0.00472
Unspecified chronic bronchitis	117 (0.261)	251 (0.366)	0.0188	96 (0.254)	99 (0.262)	0.00156
Emphysema	496 (1.11)	1213 (1.77)	0.0556	455 (1.2)	469 (1.24)	0.00337
Other chronic obstructive pulmonary disease	1458 (3.26)	3218 (4.7)	0.0737	1310 (3.46)	1352 (3.57)	0.00602

Asthma	3568 (7.97)	4909 (7.16)	0.0305	2869 (7.58)	2841 (7.51)	0.0028
Bronchiectasis	157 (0.351)	355 (0.518)	0.0254	149 (0.394)	136 (0.359)	0.00561
Alcoholic liver disease	135 (0.301)	392 (0.572)	0.041	131 (0.346)	131 (0.346)	0
Hepatic failure; not elsewhere classified	196 (0.438)	320 (0.467)	0.00434	161 (0.425)	170 (0.449)	0.0036
Chronic hepatitis; not elsewhere classified	34 (0.076)	61 (0.089)	0.00455	30 (0.079)	29 (0.077)	0.000947
Fibrosis and cirrhosis of liver	358 (0.799)	706 (1.03)	0.0242	317 (0.838)	346 (0.914)	0.00822
Fatty (change of) liver; not elsewhere classified	1000 (2.23)	1304 (1.9)	0.0233	781 (2.06)	810 (2.14)	0.00534
Chronic passive congestion of liver	168 (0.375)	360 (0.525)	0.0224	149 (0.394)	146 (0.386)	0.00127
Portal hypertension	131 (0.293)	309 (0.451)	0.026	123 (0.325)	139 (0.367)	0.0072
Other specified diseases of liver	621 (1.39)	1195 (1.74)	0.0287	539 (1.42)	565 (1.49)	0.00573
Psoriasis	317 (0.708)	618 (0.902)	0.0217	279 (0.737)	285 (0.753)	0.00184
Rheumatoid arthritis with rheumatoid factor	154 (0.344)	377 (0.55)	0.0309	142 (0.375)	136 (0.359)	0.00262
Other rheumatoid arthritis	439 (0.98)	1189 (1.74)	0.0652	412 (1.09)	430 (1.14)	0.00454
Systemic lupus erythematosus (SLE)	202 (0.451)	243 (0.355)	0.0153	155 (0.41)	157 (0.415)	0.000825
Chronic kidney disease (CKD)	2656 (5.93)	4533 (6.61)	0.0281	2238 (5.91)	2215 (5.85)	0.00258
Renal Transplantation Procedures	65 (0.145)	49 (0.071)	0.0224	39 (0.103)	35 (0.092)	0.00338
Liver Transplantation Procedures	19 (0.042)	23 (0.034)	0.00456	16 (0.042)	13 (0.034)	0.00405

SMD=Standardized Mean Difference, SBP=Systolic Blood Pressure, DBP=Diastolic Blood Pressure, BMI=Body Mass Index

Table 7 – Characteristics of the cohorts with and without a diagnosis of a psychiatric illness in the past year

	Before matching			After matching		
	Psychiatric history	**No psychiatric history**	**SMD**	**Psychiatric history**	**No psychiatric history**	**SMD**
Number	1,801,488	10,818,875	-	1,729,837	1,729,837	-
Age; mean (SD); y	50.5 (18.4)	51.4 (19.1)	0.0474	50.5 (18.4)	50.5 (18.4)	0.00125
18-30	318279 (17.7)	2010738 (18.6)	0.0238	312026 (18)	312026 (18)	0
31-45	427263 (23.7)	2346582 (21.7)	0.0484	408036 (23.6)	408036 (23.6)	0
46-60	466599 (25.9)	2557409 (23.6)	0.0524	439808 (25.4)	439808 (25.4)	0
61-75	421808 (23.4)	2643211 (24.4)	0.0238	404952 (23.4)	404952 (23.4)	0
76+	167539 (9.3)	1260935 (11.7)	0.0769	165015 (9.54)	165015 (9.54)	0
Female	1206345 (67)	5999463 (55.5)	0.238	1140693 (65.9)	1140693 (65.9)	0
Male	595143 (33)	4819412 (44.5)	0.238	589144 (34.1)	589144 (34.1)	0
Race						
White	1388549 (77.1)	6640022 (61.4)	0.345	1325799 (76.6)	1329165 (76.8)	0.00461
Black or African American	203978 (11.3)	1419381 (13.1)	0.0549	196875 (11.4)	206695 (11.9)	0.0177
Asian	24572 (1.36)	398724 (3.69)	0.148	24523 (1.42)	23228 (1.34)	0.00642
American Indian or Alaska Native	6610 (0.367)	41309 (0.382)	0.00244	6306 (0.365)	5305 (0.307)	0.01
Native Hawaiian or Other Pacific Islander	1780 (0.0988)	14521 (0.134)	0.0104	1750 (0.101)	1728 (0.0999)	0.000401
Unknown	175999 (9.77)	2304918 (21.3)	0.323	174584 (10.1)	163716 (9.46)	0.0212
Blood Pressure; Systolic						
At most 140 mm[Hg]	1146811	3918621	0.571	1079277	1090062	0.0129
At least 160 mm[Hg]	364902	984082	0.319	327269	324536	0.00404
140-160 mm[Hg]	723341	2138029	0.457	662935	667235	0.00511
Blood Pressure; Diastolic						
At most 90 mm[Hg]	1162693	4060149	0.561	1095111	1106739	0.014
90-100 mm[Hg]	583018	1472539	0.457	525523	524474	0.00132
At least 100 mm[Hg]	278291	588045	0.332	240926	235288	0.00946
Hypertensive diseases (deprecated 2018)	667223 (37)	2041125 (18.9)	0.413	613047 (35.4)	626416 (36.2)	0.0161
Hypertensive diseases	667465 (37.1)	2041990 (18.9)	0.414	613277 (35.5)	626551 (36.2)	0.016
Hypertensive chronic kidney disease	66803 (3.71)	152584 (1.41)	0.146	58948 (3.41)	55857 (3.23)	0.00998
BMI						
At most 25 kg/m2	420085	1299496	0.3	395101	391615	0.00481
25-30 kg/m2	485438	1470669	0.337	450528	451455	0.00122
At least 30 kg/m2	508343	1396556	0.386	461655	470564	0.0116
Overweight and obesity	404564 (22.5)	832859 (7.7)	0.422	355375 (20.5)	358856 (20.7)	0.00497
Type 1 diabetes mellitus	48981 (2.72)	108731 (1.01)	0.127	41220 (2.38)	37387 (2.16)	0.0149
Type 2 diabetes mellitus	273154 (15.2)	817561 (7.56)	0.241	245369 (14.2)	242117 (14)	0.0054
Nicotine dependence	296383 (16.5)	483069 (4.47)	0.399	247384 (14.3)	244857 (14.2)	0.00418
Ischemic heart diseases	184694 (10.3)	631158 (5.83)	0.163	165508 (9.57)	159459 (9.22)	0.012
Other forms of heart disease	358098 (19.9)	1015125 (9.38)	0.3	319250 (18.5)	312426 (18.1)	0.0102
Bronchitis; not specified as acute or chronic	116576 (6.47)	184226 (1.7)	0.243	92699 (5.36)	90600 (5.24)	0.00542
Simple and mucopurulent chronic bronchitis	13986 (0.776)	14595 (0.135)	0.0954	9366 (0.541)	8107 (0.469)	0.0103
Unspecified chronic bronchitis	15367 (0.853)	15835 (0.146)	0.1	10280 (0.594)	8771 (0.507)	0.0118
Emphysema	41744 (2.32)	63149 (0.584)	0.145	32084 (1.85)	29023 (1.68)	0.0134
Other chronic obstructive pulmonary disease	128244 (7.12)	205367 (1.9)	0.254	100283 (5.8)	93848 (5.43)	0.0162
Asthma	246826 (13.7)	467406 (4.32)	0.332	207016 (12)	202661 (11.7)	0.00779
Bronchiectasis	11122 (0.617)	23683 (0.219)	0.0618	9257 (0.535)	8121 (0.469)	0.00929
Chronic kidney disease (CKD)	119993 (6.66)	318544 (2.94)	0.174	107295 (6.2)	102778 (5.94)	0.0109

Table 8 – Hazard ratios of different psychiatric sequelae between COVID-19 and the 6 control health events

	Influenza	**Other RTI**	**Skin infection**	**Cholelithiasis**	**Urolithiasis**	**Fracture**
	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)
Psychiatric illness	2.08 (1.77-2.46)	1.67 (1.47-1.90)	1.63 (1.42-1.87)	1.58 (1.32-1.88)	2.24 (1.90-2.65)	2.14 (1.86-2.47)
Psychotic disorder	3.43 (1.30-9.03)	1.52 (0.74-3.12)	0.87 (0.43-1.75)	1.63 (0.64-4.15)	4.02 (1.65-9.81)	0.92 (0.48-1.76)
Mood disorder	1.79 (1.37-2.33)	1.33 (1.09-1.63)	1.06 (0.86-1.31)	1.22 (0.93-1.59)	1.62 (1.26-2.07)	1.35 (1.09-1.67)
Mania/Bipolar	2.24 (0.91-5.50)	0.51 (0.24-1.08)	0.47 (0.22-0.98)	2.02 (0.75-5.43)	0.85 (0.36-2.03)	0.52 (0.25-1.09)
Depressive episode	2.07 (1.55-2.78)	1.61 (1.28-2.03)	1.20 (0.95-1.51)	1.26 (0.95-1.69)	1.66 (1.26-2.17)	1.29 (1.03-1.63)
Persistent mood disorder	1.24 (0.50-3.07)	1.49 (0.60-3.66)	1.21 (0.49-3.01)	1.57 (0.47-5.28)	3.18 (1.12-9.03)	2.96 (0.98-8.95)
Anxiety disorder	2.25 (1.87-2.69)	1.89 (1.63-2.18)	1.96 (1.68-2.29)	1.59 (1.30-1.94)	2.51 (2.08-3.03)	2.62 (2.22-3.09)
Phobia	7.19 (2.10-24.61)	1.85 (0.72-4.73)	0.92 (0.36-2.34)	0.93 (0.35-2.44)	1.41 (0.57-3.50)	2.27 (0.80-6.47)
Other anxiety disorder	2.19 (1.79-2.67)	1.81 (1.54-2.13)	1.84 (1.55-2.18)	1.42 (1.13-1.77)	2.09 (1.70-2.56)	2.68 (2.22-3.24)
OCD	5.79 (1.26-26.62)	1.56 (0.56-4.33)	0.51 (0.20-1.29)	-	4.45 (1.22-16.27)	3.09 (0.96-9.91)
Reaction to severe stress	2.81 (1.98-3.99)	2.40 (1.81-3.17)	2.47 (1.83-3.34)	2.35 (1.60-3.44)	4.63 (3.07-6.97)	2.50 (1.86-3.37)
Dissociative disorder	0.69 (0.15-3.08)	1.50 (0.58-3.86)	3.08 (0.92-10.28)	-	5.10 (1.10-23.64)	1.99 (0.59-6.73)
Somatoform disorder	2.91 (1.02-8.29)	1.42 (0.57-3.50)	2.13 (0.84-5.42)	5.79 (1.14-29.43)	-	6.26 (1.68-23.37)
Other neurotic disorder	3.30 (0.69-15.90)	1.56 (0.45-5.34)	0.89 (0.30-2.69)	-	-	3.70 (0.77-17.85)
Insomnia	3.26 (2.36-4.49)	2.35 (1.84-3.01)	2.15 (1.67-2.77)	1.85 (1.35-2.54)	3.29 (2.38-4.53)	2.43 (1.88-3.15)
Dementia	3.13 (1.57-6.25)	2.05 (1.33-3.15)	1.40 (0.90-2.19)	3.01 (1.72-5.29)	4.43 (2.44-8.02)	1.38 (0.92-2.08)
Dementia (among 65+)	2.08 (1.77-2.46)	1.67 (1.47-1.90)	1.63 (1.42-1.87)	1.58 (1.32-1.88)	2.24 (1.90-2.65)	2.14 (1.86-2.47)

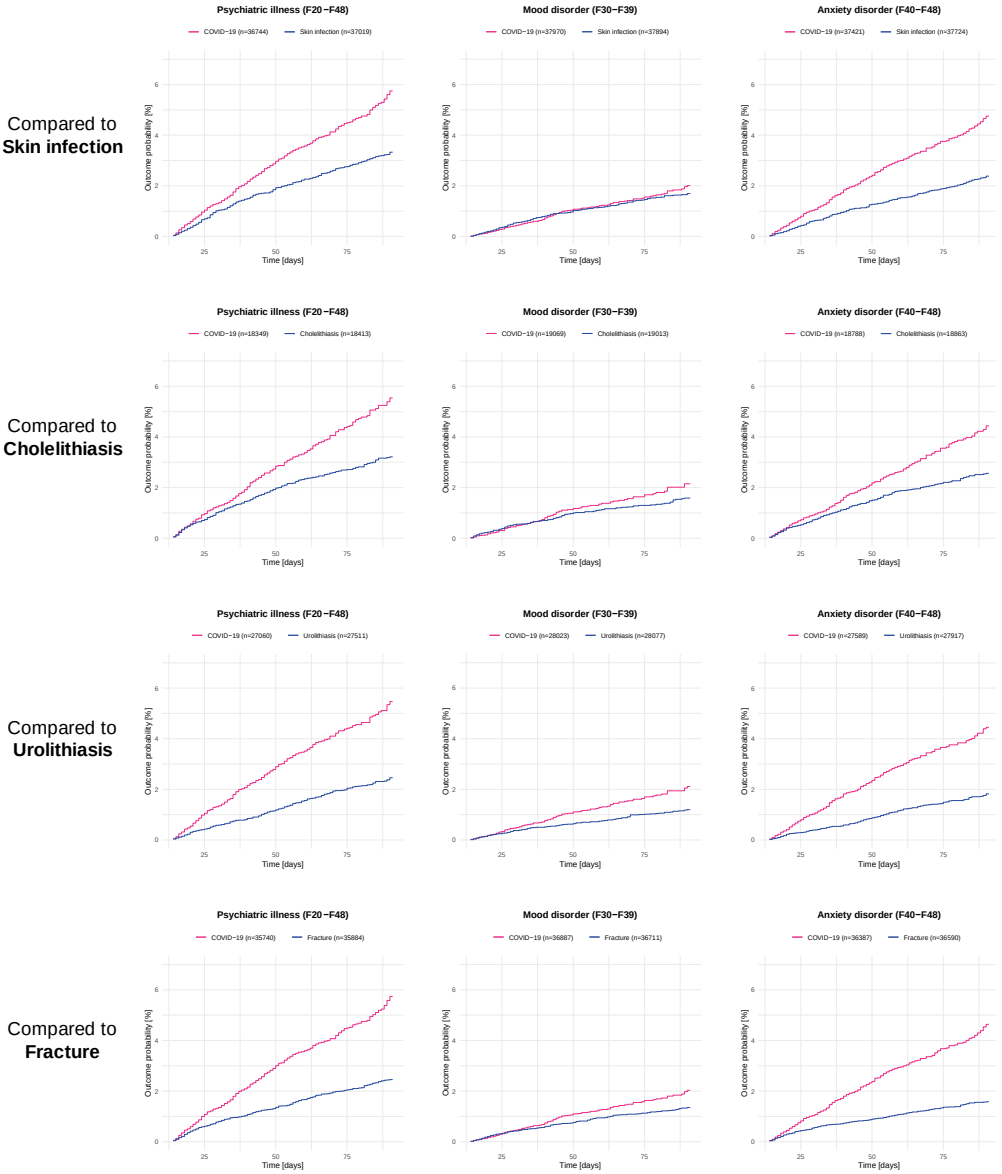

Fig. 1 – Kaplan-Meier curves representing the onset of first psychiatric diagnoses after COVID-19 compared to the other 4 control health events not presented in Fig. 1 of the main manuscript. Shaded areas represent 95% confidence intervals.

Hazard ratio over time

The p-values for the test of proportional hazard of the original model, the hazard ratios for the first and second phases of the follow-up period, and the p-values for the test of proportional hazard of the model with follow-up split in two phases are summarised in Table 9 below. Comparison with three control health events showed evidence of non-proportionality of the hazard: skin infections (proportional hazard test: p=0.042), cholelithiasis (p=0.0082), and fracture (p=0.0030). In the comparison to these three cohorts, the hazard ratio tended to increase between the early phase and the late phase. However, the hazard ratios remained statistically significant for both phases of the follow-up period (all p<0.0001 except for cholelithiasis in the early phase: p=0.0044).

It is important to note that our detection of non-proportionality comes in part from the ability to detect small effects with large sample sizes. For instance, in the comparison with the fracture cohort, decreasing the sample size by removing at random 80% of the subjects leads to over 75% of the models showing no sign of non-proportionality while 100% of them still show a significant HR of the same value on average as the one calculated using the whole sample (data not shown). Consequently, we elected to report the hazard ratios for all 6 comparison cohorts. We also report the total probability of outcome at 90 days (Table 2 of the manuscript) which does not depend on proportional hazard assumption.

Since the COVID-19 cohort remains the same across all 6 comparisons, differences in non-proportionality must come from the control cohorts. As can be observed in Fig. 1 of the appendix (see above), the Kaplan-Meier curves for the skin infection, cholelithiasis, and fracture cohorts appear to flatten relatively to the curve for the COVID-19 cohort. This helps explain why the hazard ratio increases in the second part of the follow-up period.

When evaluating a piecewise constant hazard ratio, there was no longer evidence for non-proportionality in the skin infection and cholelithiasis cohorts (p>0.05) and little evidence for non-proportionality in the fracture cohort (p=0.010). Plotting the scaled Schoenfeld residuals against transformed time reveals why that is: even within the early phase, the hazard ratio tends to increase with time. Further splitting this epoch in two (14-28 days and 29-44 days) leads to HR that are monotonically increasing: from 1.67 (95% CI: 1.32–2.09, p<0.0001) in the 14-28 days period, to 2.44 (95% CI: 1.84–3.24, p<0.0001) in the 29-44 days period, to 2.56 (95% CI: 2.02–3.24, p<0.0001) in the late phase. When accounting for this additional splitting of the time interval, there was no longer evidence of non-proportionality (p=0.12).

In summary, the risk of a first diagnosis of psychiatric illness was significantly greater after COVID-19 than after all 6 control health events and this was the case both during the early and late phases of the follow up period. However, for 3 of the control health events, the risk started to plateau in the late phase of the follow-up period whereas there was no sign of plateauing at 90 days after COVID-19. As a result, in the comparisons to these 3 control health events, the hazard ratio was higher in the late phase of the follow-up than in the early phase.

Table 9 – Hazard ratios for a first diagnosis of a psychiatric illness (F20-F48), test of proportional hazards, and hazard ratios for the split follow-up

	Whole follow-up		Early/Late phases		
		PH test	14-44 days	45-90 days	PH test
	HR (95% CI)	p-value	HR (95% CI)	HR (95% CI)	p-value
Influenza	2.1 (1.8-2.5)	0.75	2.1 (1.7-2.6)	2.1 (1.6-2.7)	0.72
Other RTI	1.7 (1.5-1.9)	0.78	1.6 (1.4-1.9)	1.7 (1.4-2.1)	0.99
Skin infection	1.6 (1.4-1.9)	0.042	1.5 (1.2-1.7)	2.0 (1.6-2.5)	0.93
Cholelithiasis	1.6 (1.3-1.9)	0.0082	1.4 (1.1-1.7)	2.0 (1.5-2.7)	0.18
Urolithiasis	2.2 (1.9-2.6)	0.85	2.4 (1.9-3.0)	2.0 (1.6-2.6)	0.38
Fracture	2.1 (1.9-2.5)	0.0030	1.9 (1.6-2.3)	2.6 (2.0-3.2)	0.010

PH test p-value corresponds to the test for proportional hazard and a small value indicates evidence for non-proportional hazard.

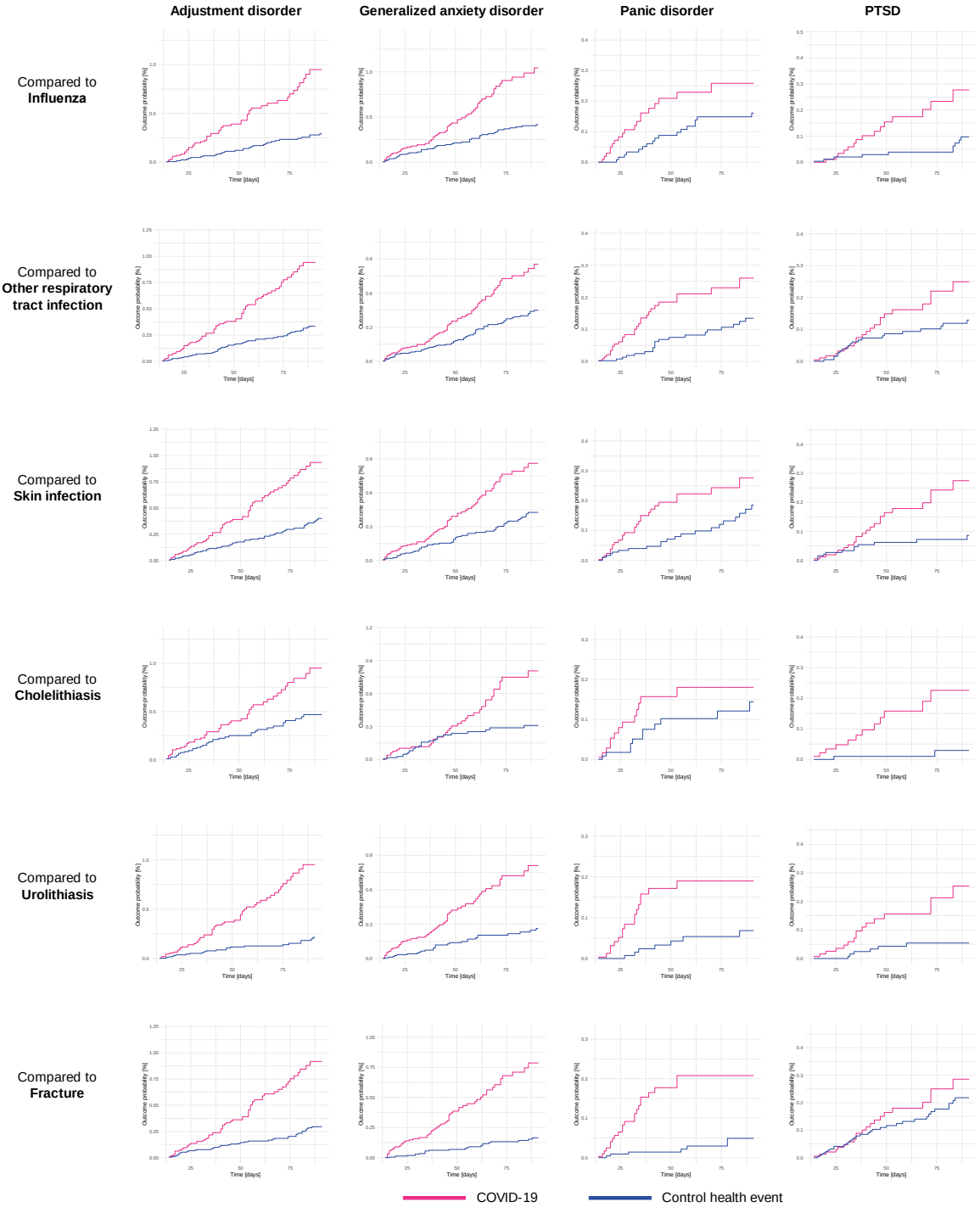

Fig. 2 – Kaplan-Meier curves for the most common sequelae of COVID-19 in terms of anxiety disorder subcategories. Shaded areas represent 95% confidence intervals.

Table 10 – Probability of a first diagnosis of anxiety disorders after COVID-19 and control health events within the period 14 days to 90 days post-diagnosis. P-values are from logrank tests.

	COVID-19	Influenza		Other RTI		Skin infection	
	% (95% CI)	% (95% CI)	p	% (95% CI)	p	% (95% CI)	p
Phobia	0.11 (0.05-0.24)	0.02 (0.005-0.08)	0.0017	0.04 (0.02-0.10)	0.091	0.075 (0.04-0.14)	0.69
Other anxiety disorder	3.7 (3.2-4.2)	1.8 (1.6-2.1)	<0.0001	2 (1.8-2.3)	<0.0001	2 (1.7-2.2)	<0.0001
Panic disorder	0.26 (0.17-0.4)	0.16 (0.1-0.26)	0.058	0.13 (0.09-0.21)	0.0045	0.19 (0.12-0.29)	0.018
GAD	0.85 (0.66-1.1)	0.42 (0.31-0.56)	<0.0001	0.45 (0.35-0.57)	0.0003	0.43 (0.32-0.57)	0.00013
OCD	0.04 (0.008-0.17)	0.009 (0.001-0.07)	0.18	0.019 (0.006-0.06)	0.83	0.06 (0.03-0.13)	0.22
Reac. to severe stress	1.3 (1.1-1.7)	0.5 (0.38-0.66)	<0.0001	0.58 (0.47-0.72)	<0.0001	0.55 (0.43-0.71)	<0.0001
Acute stress reaction	0.21 (0.13-0.34)	0.053 (0.024-0.12)	0.0025	0.065 (0.034-0.12)	0.0015	0.05 (0.02-0.13)	<0.0001
PTSD	0.25 (0.16-0.4)	0.097 (0.051-0.18)	0.0046	0.13 (0.082-0.2)	0.038	0.09 (0.05-0.16)	0.0023
Adjustment disorders	0.94 (0.74-1.2)	0.29 (0.2-0.42)	<0.0001	0.34 (0.25-0.45)	<0.0001	0.4 (0.3-0.54)	<0.0001
Dissociative disorder	0.05 (0.02-0.17)	0.017 (0.0042-0.067)	0.77	0.03 (0.01-0.07)	0.68	0.01 (0.003-0.05)	0.17
Somatoform disorder	0.12 (0.05-0.28)	0.03 (0.0096-0.092)	0.068	0.04 (0.02-0.09)	0.18	0.05 (0.02-0.12)	0.20
Oth. neurotic disorder	0.02 (0.004-0.07)	0.008 (0.001-0.06)	0.35	0.01 (0.003-0.05)	0.68	0.03 (0.009-0.09)	0.95

	Cholelithiasis		Urolithiasis		Fracture	
	% (95% CI)	p	% (95% CI)	p	% (95% CI)	p
Phobia	0.09 (0.04-0.2)	0.78	0.06 (0.03-0.13)	0.38	0.04 (0.02-0.09)	0.13
Other anxiety disorder	2.1 (1.8-2.5)	0.0021	1.7 (1.4-2)	<0.0001	1.2 (1-1.4)	<0.0001
Panic disorder	0.14 (0.08-0.27)	0.21	0.07 (0.03-0.15)	0.0018	0.05 (0.02-0.11)	<0.0001
GAD	0.31 (0.21-0.46)	0.0028	0.26 (0.18-0.39)	<0.0001	0.16 (0.11-0.25)	<0.0001
OCD	0 (-)	0.23	0.01 (0.002-0.08)	0.22	0.01 (0.003-0.05)	0.41
Reac. to severe stress	0.56 (0.41-0.76)	<0.0001	0.3 (0.21-0.44)	<0.0001	0.52 (0.41-0.66)	<0.0001
Acute stress reaction	0.06 (0.02-0.15)	0.029	0 (-)	<0.0001	0.03 (0.01-0.07)	<0.0001
PTSD	0.03 (0.007-0.13)	0.00039	0.05 (0.02-0.12)	0.00058	0.22 (0.15-0.31)	0.29
Adjustment disorders	0.47 (0.33-0.66)	0.0037	0.21 (0.14-0.33)	<0.0001	0.3 (0.22-0.40)	<0.0001
Dissociative disorder	0 (-)	0.095	0.007 (0.001-0.05)	0.16	0.02 (0.007-0.05)	0.34
Somatoform disorder	0.01 (0.002-0.10)	0.098	0 (-)	0.002	0.02 (0.006-0.06)	0.0077
Oth. neurotic disorder	0 (-)	1	0 (-)	1	0.008 (0.001-0.05)	0.22

GAD=Generalized anxiety disorder, OCD=Obsessive-compulsive disorder, PTSD=Post-traumatic stress disorder

Table 11 – Probability of a first diagnosis of mood disorder after COVID-19 and control health events within the period 14 days to 90 days post-diagnosis. P-values are from logrank tests.

	COVID-19	Influenza		Other RTI		Skin infection	
	% (95% CI)	% (95% CI)	p	% (95% CI)	p	% (95% CI)	p
Mania/Bipolar (F30-F31)	0.1 (0.04-0.25)	0.06 (0.03-0.15)	0.15	0.16 (0.11-0.24)	0.17	0.23 (0.16-0.35)	0.083
Depressive episode (F32)	1.7 (1.4-2.1)	0.79 (0.64-0.98)	<0.0001	1.0 (0.89-1.2)	<0.0001	1.2 (1.0-1.4)	0.12

	Cholelithiasis		Urolithiasis		Fracture	
	% (95% CI)	p	% (95% CI)	p	% (95% CI)	p
Mania/Bipolar (F30-F31)	0.05 (0.02-0.16)	0.15	0.10 (0.06-0.19)	0.90	0.17 (0.11-0.25)	0.24
Depressive episode (F32)	1.3 (1.0-1.5)	0.11	1.0 (0.82-1.2)	0.00021	1.1 (0.97-1.3)	0.027

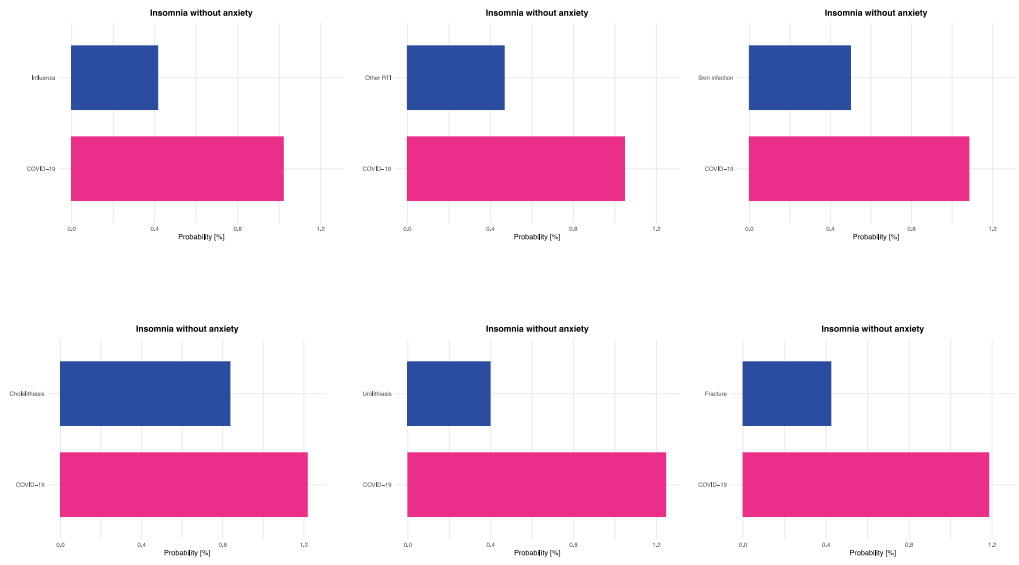

Fig. 3 – Probability of insomnia in the absence of a concurrent diagnosis of anxiety disorder within the 90 days following a diagnosis of COVID-19 compared to control health events

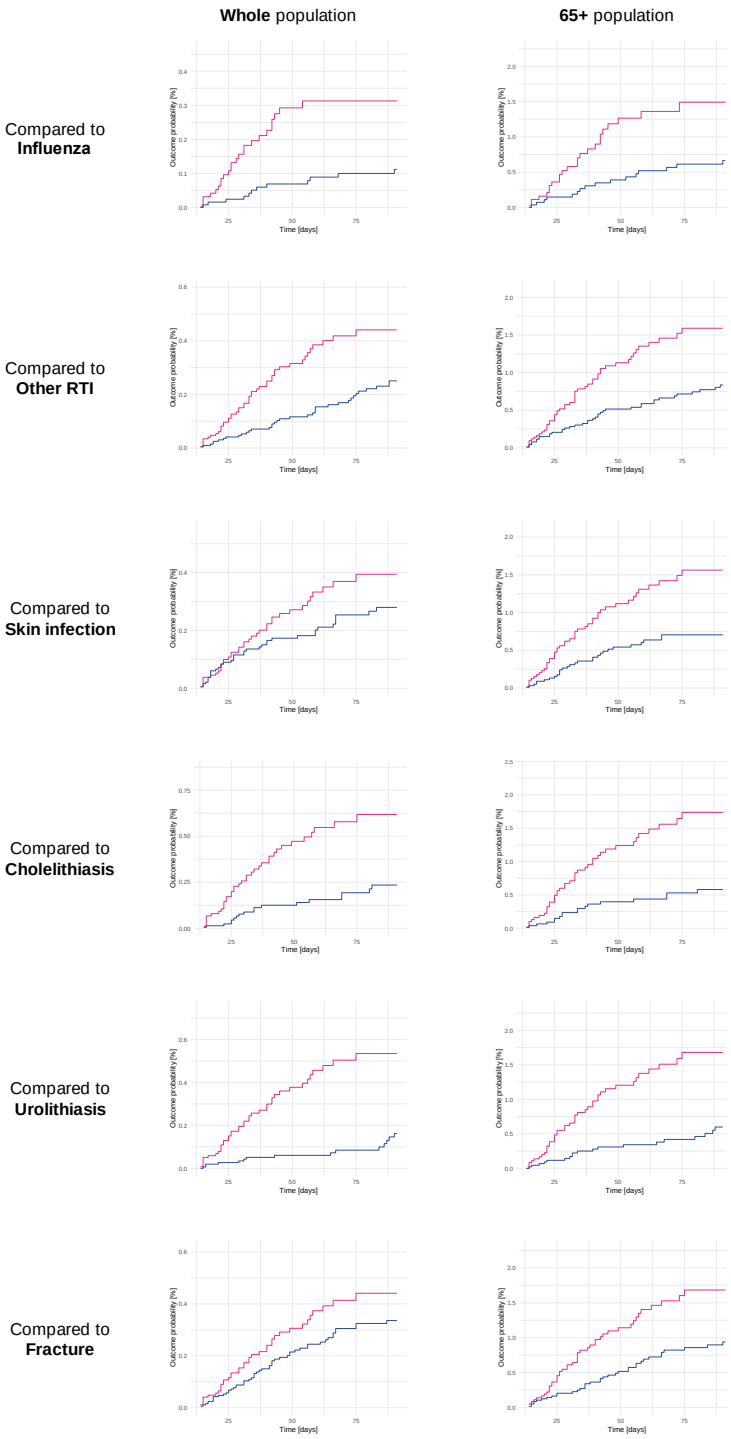

Fig. 4 – Kaplan-Meier curves for the diagnosis of dementia after COVID-19 (red) and other health events (blue) in the whole population (left) and within patients over the age of 65 (right). Shaded areas represent 95% confidence intervals.

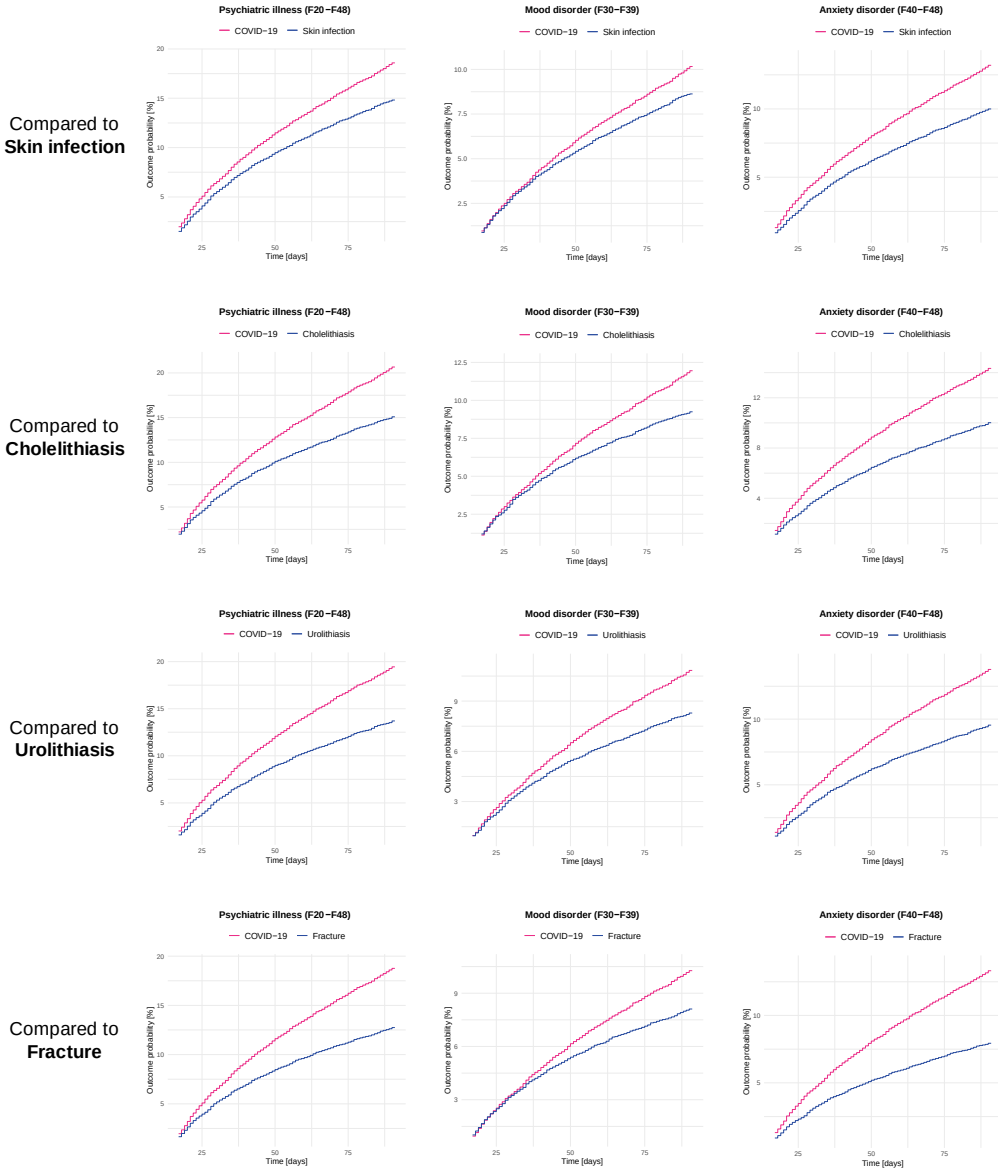

Fig. 5 – Kaplan-Meier curves for any (first or recurrent) psychiatric diagnoses after COVID-19 compared to the other 4 control health events not presented in Fig. 2 of the main manuscript. Shaded areas represent 95% confidence intervals.

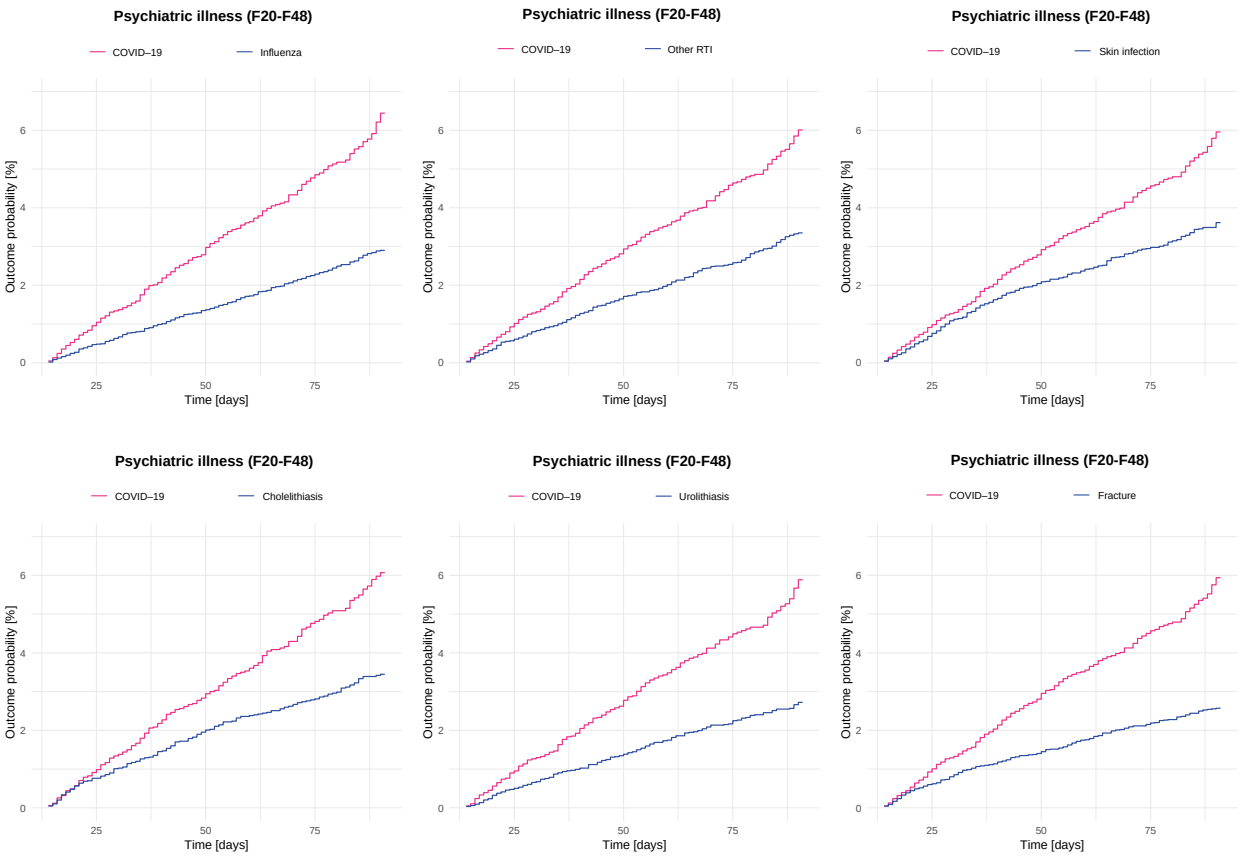

Fig. 6 – Kaplan-Meier curves for the replication of the analysis with only patients whose race is known. Shaded areas represent 95% confidence intervals.

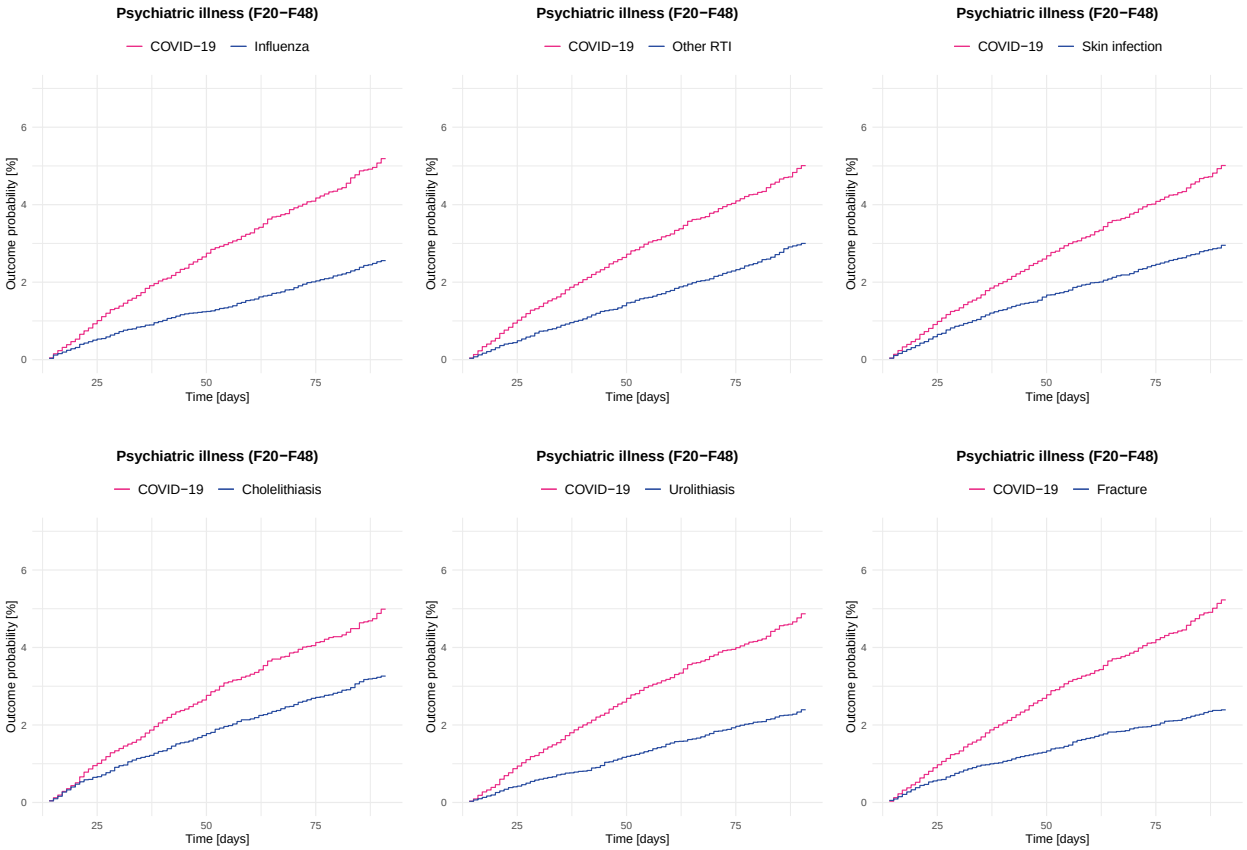

Fig. 7 – Kaplan-Meier curves for the replication of the analysis after controlling for socioeconomic factors represented by the Z59 code ('Problems related to housing and economic circumstances'). Shaded areas represent 95% confidence intervals.

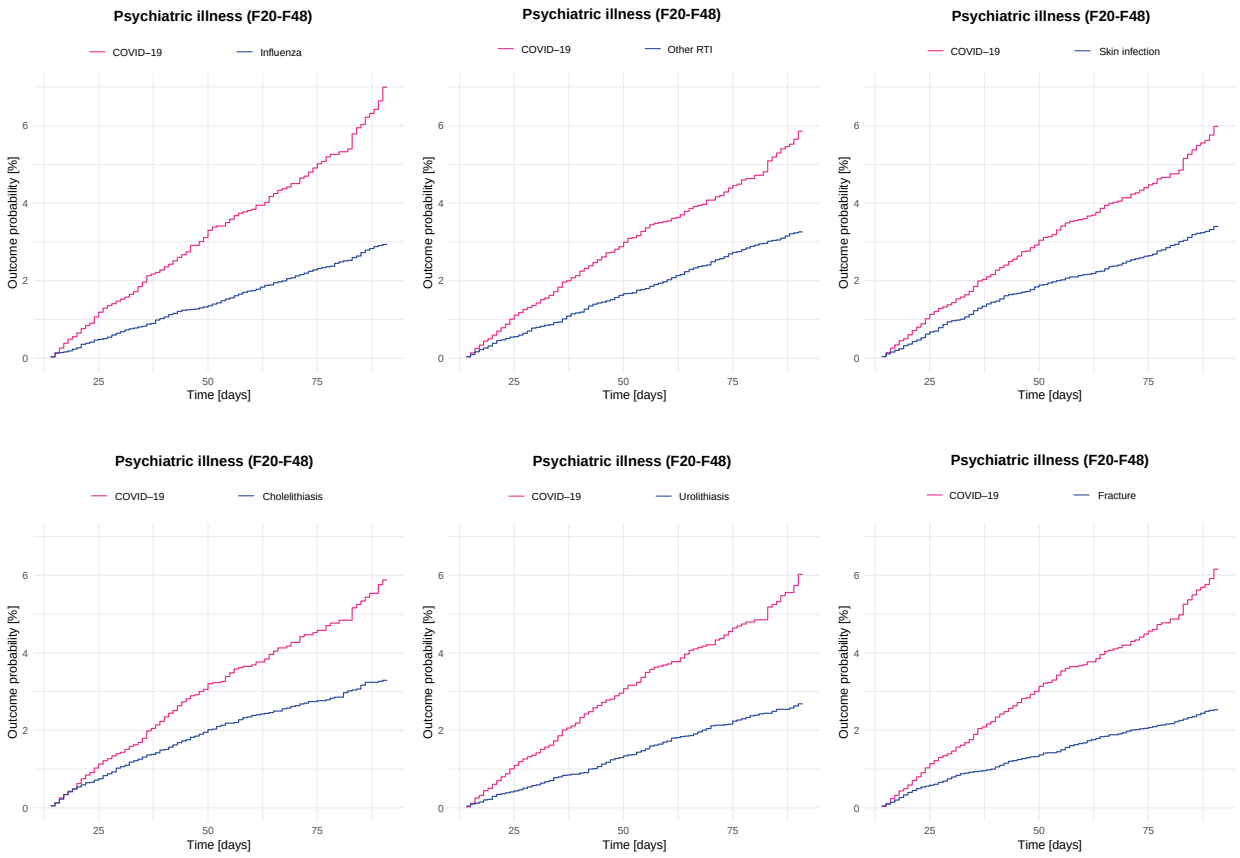

Fig. 8 – Kaplan-Meier curves for the replication of the analysis using confirmed diagnosis of COVID-19 as a stricter inclusion criterion into the cohort. Shaded areas represent 95% confidence intervals.

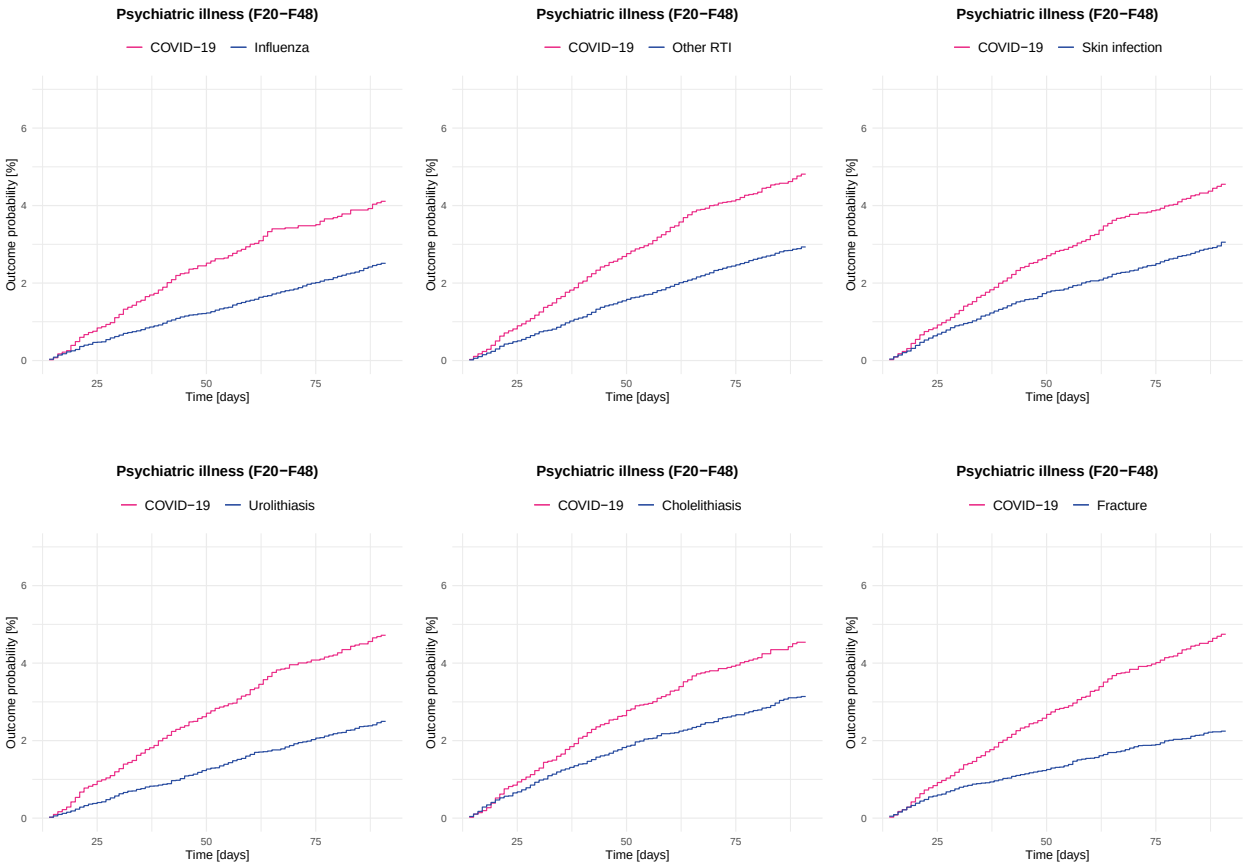

Fig. 9 – Kaplan-Meier curves for the replication of the analysis using diagnosis of COVID-19 confirmed with RNA/Antigen test as an event stricter inclusion criterion into the cohort. Shaded areas represent 95% confidence intervals.

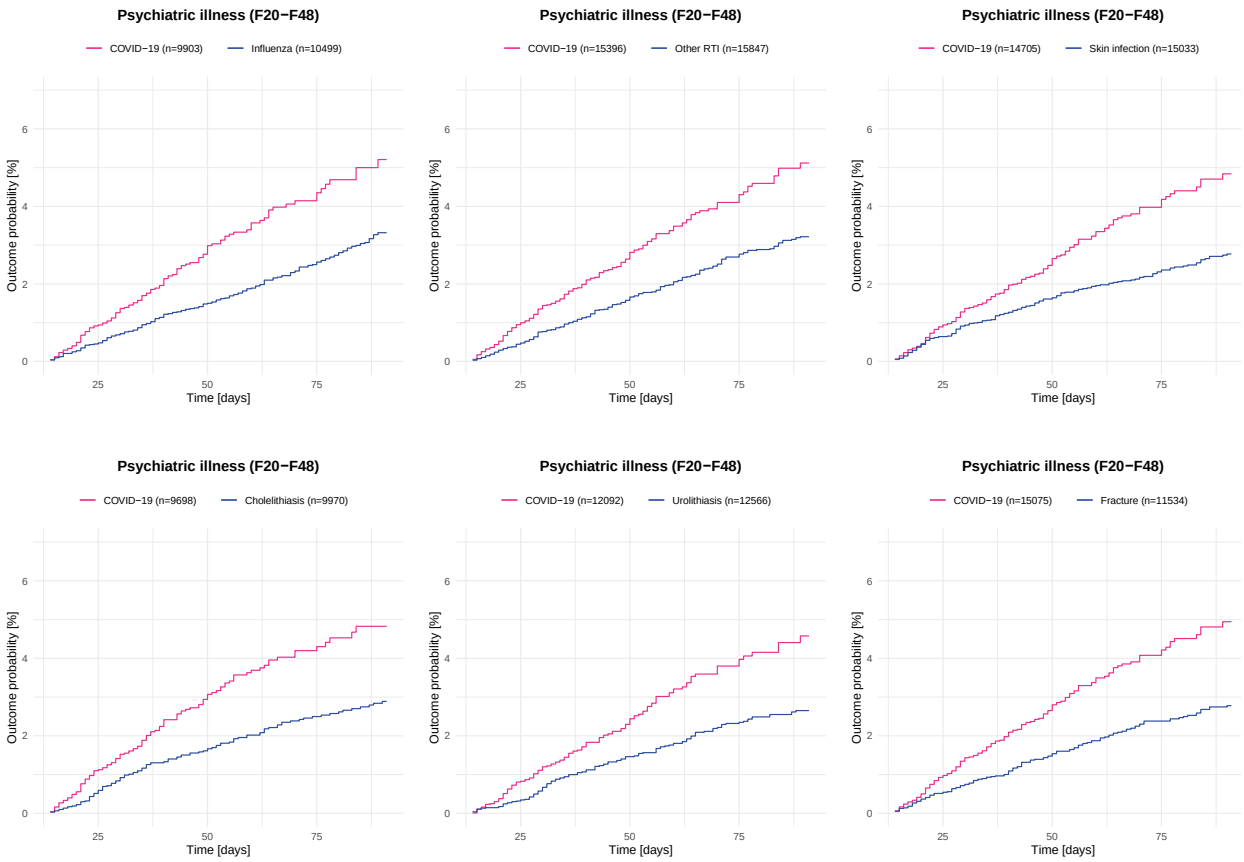

Fig. 10 – Kaplan-Meier curves for the replication of the analysis limited to patients who made at least one healthcare visit between 14 and 90 days after their diagnosis of COVID-19 (or control health event). Shaded areas represent 95% confidence intervals.

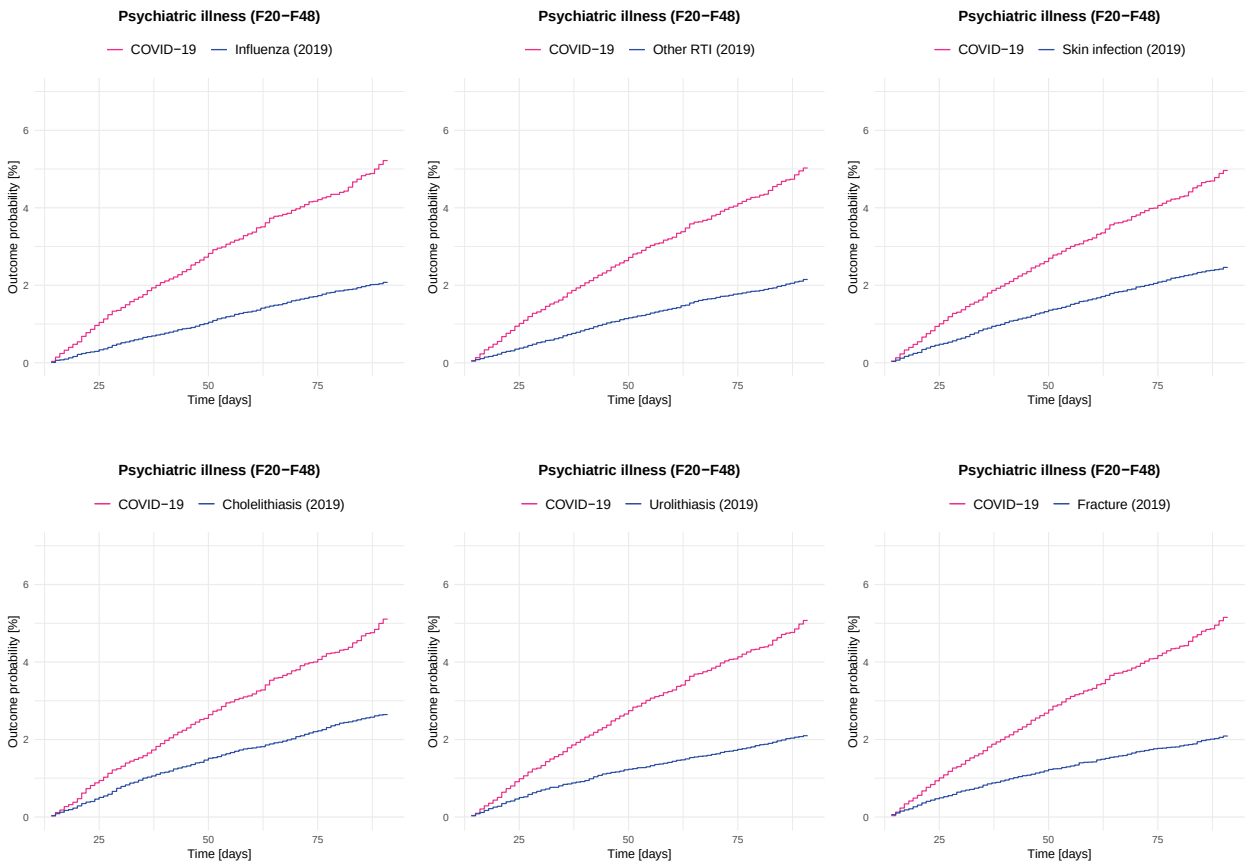

Fig. 11 – Kaplan-Meier curves for the replication of the analysis wherein the psychiatric sequelae for the control health events were measured in 2019 (before the COVID-19 pandemic). Shaded areas represent 95% confidence intervals.

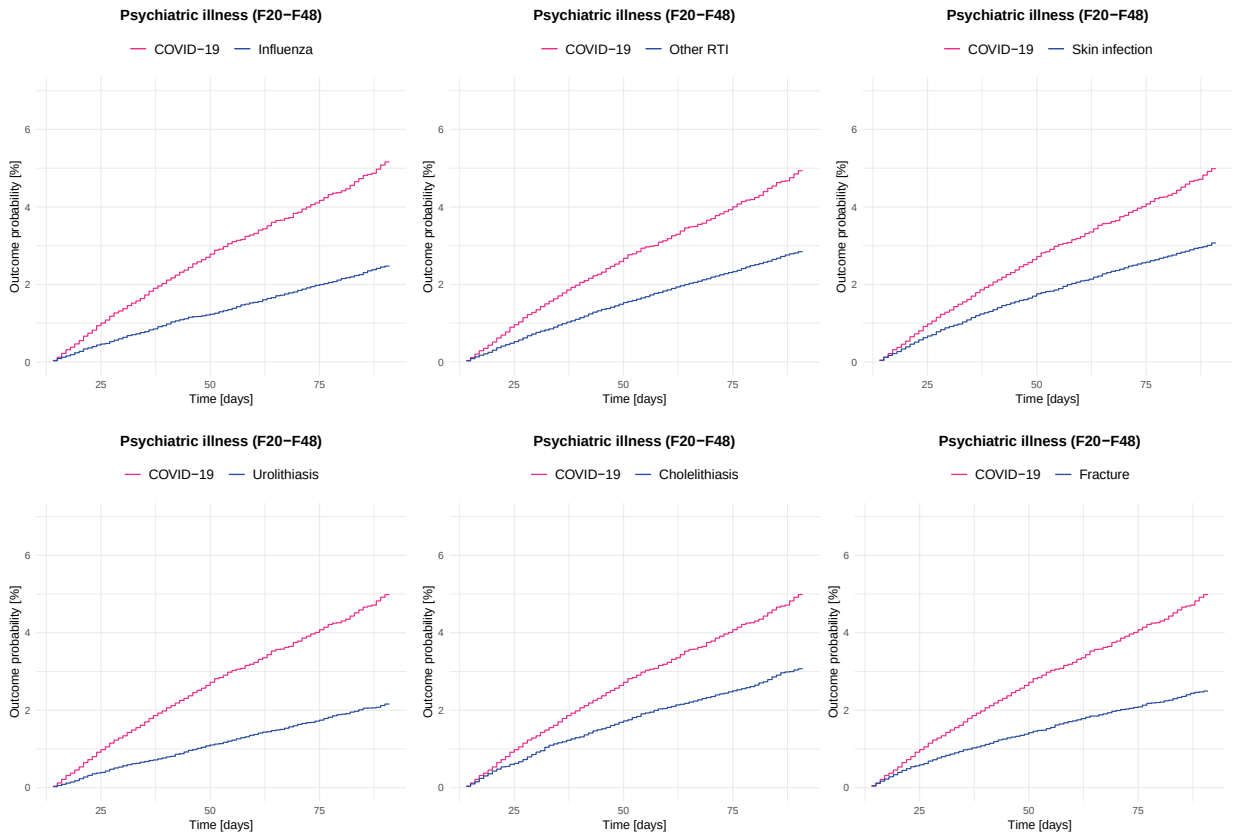

Fig. 12 – Kaplan-Meier curves for the replication of the analysis using unmatched cohorts. Shaded areas represent 95% confidence intervals.

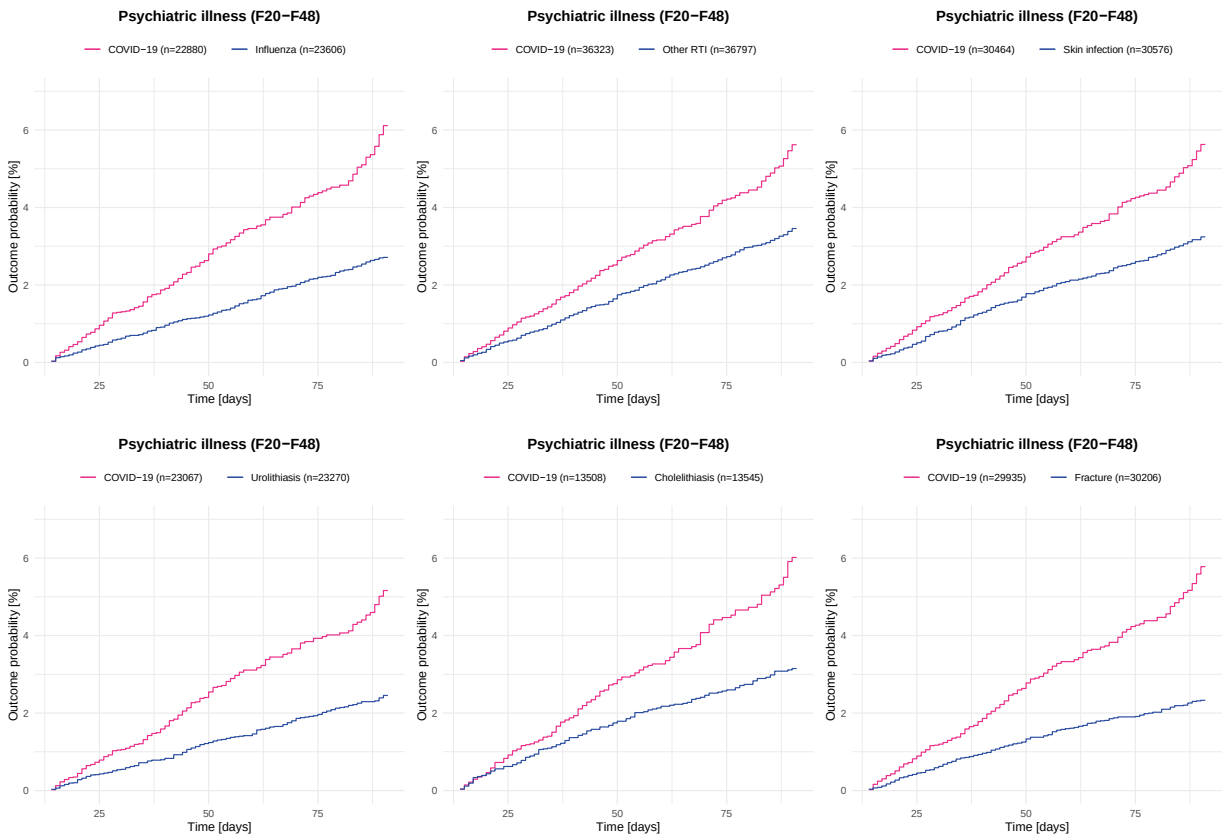

Fig. 13 – Kaplan-Meier curves for the replication of the study in patients who did not require admission to the hospital, thereby testing the "severity" hypothesis.

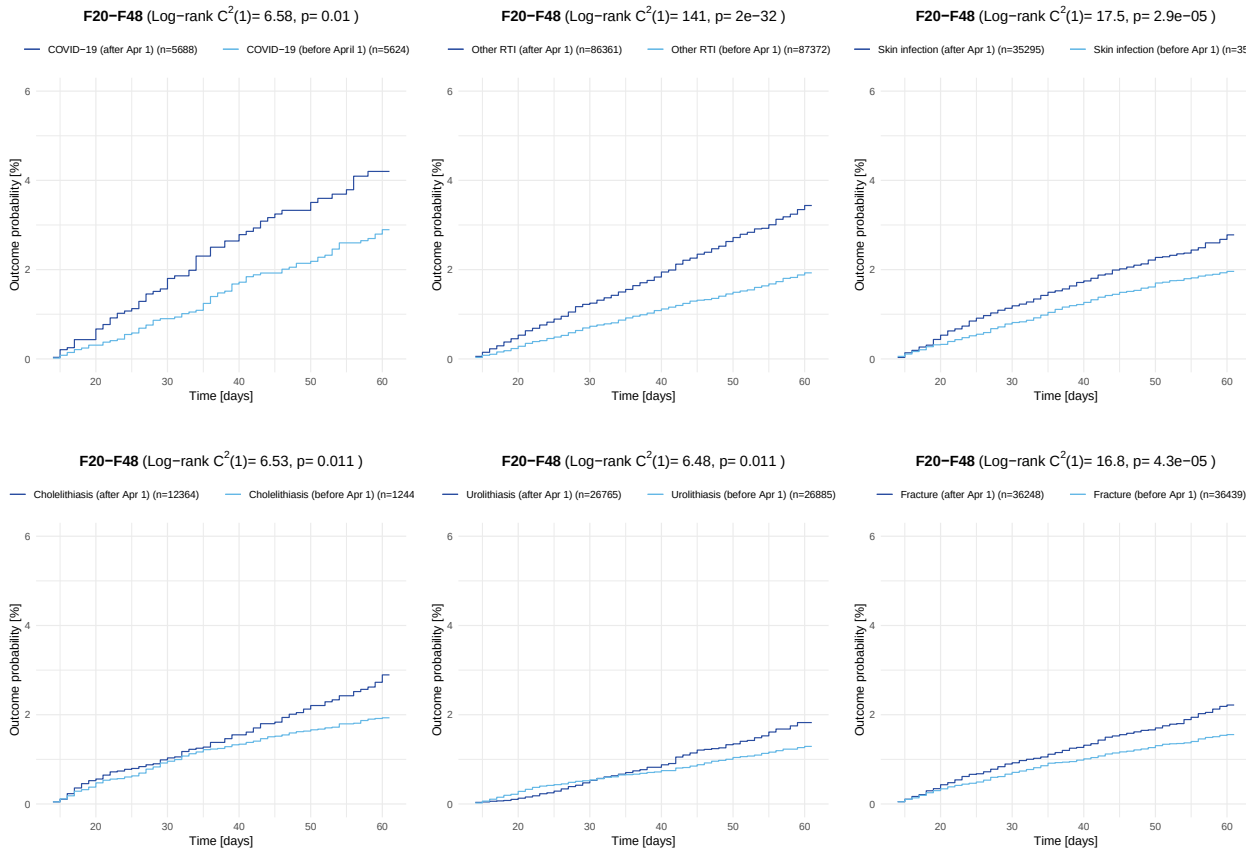

Fig. 14 – Kaplan-Meier curves comparing the incidence of psychiatric sequelae for each health event when it occurred after vs. before April 1, 2020. For all health events, the incidence of psychiatric sequelae was significant higher when the event occurred after April 1, 2020.

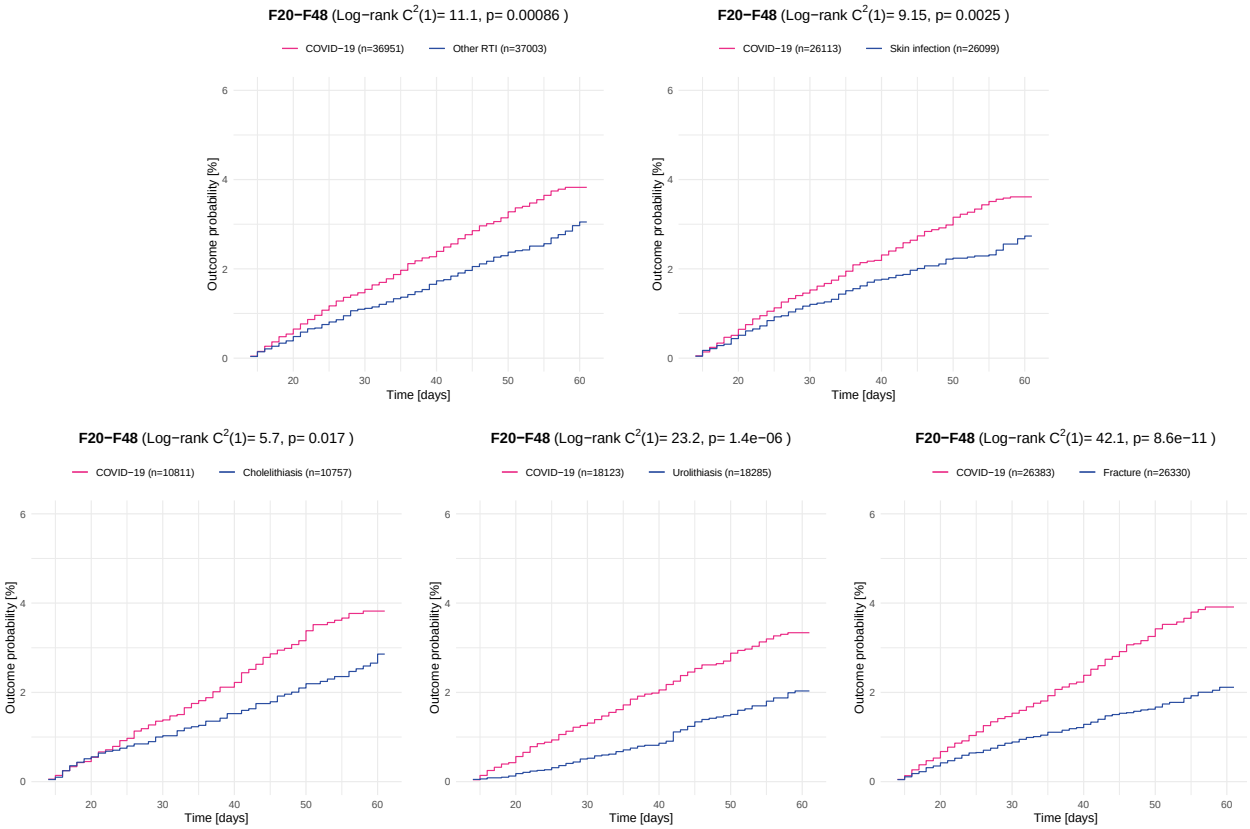

Fig. 15 – Kaplan-Meier curves comparing the incidence of psychiatric sequelae between COVID-19 and each control health event when they occurred after April 1, 2020. The contrast between groups are lower (suggesting that contextual factors play a role in the differential incidence of psychiatric sequelae between health events). However, the rate of psychiatric sequelae remains significantly larger among patients with COVID-19 suggesting a direct role for the latter in the increased rate of psychiatric sequelae.

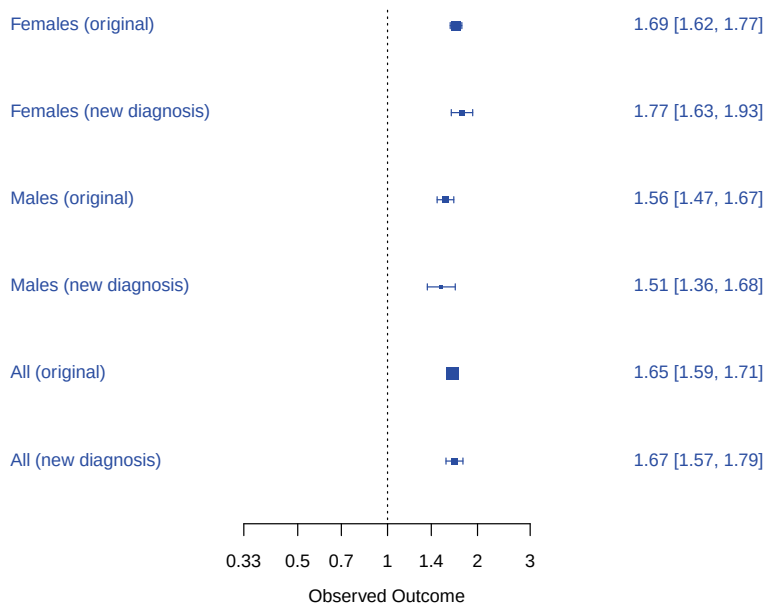

Fig. 16 – Relative risk of developing COVID-19 among patients with (vs without) a recent psychiatric history. Comparison is made between those whose recent diagnosis was a *new* diagnosis vs the original finding in which all diagnoses recorded in the past year were included in the cohort.

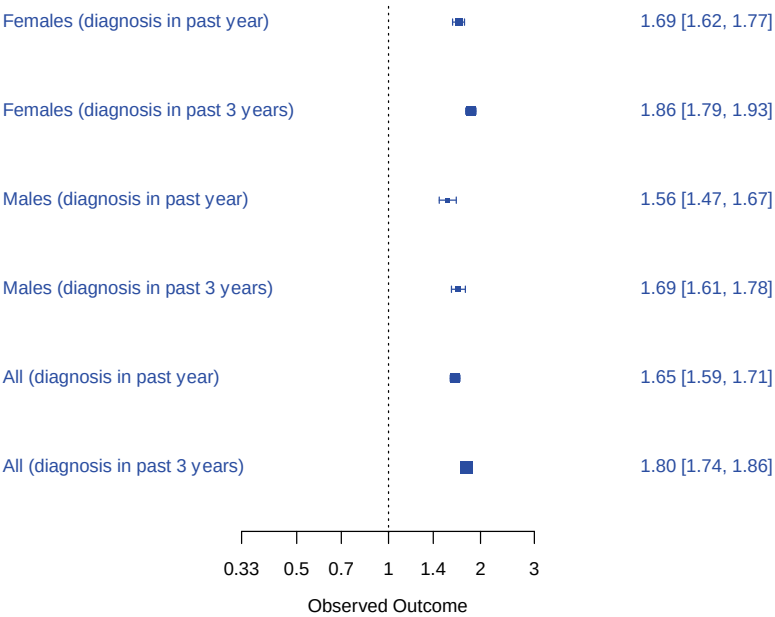

Fig. 17 – Relative risk of developing COVID-19 among patients with (vs without) a recent psychiatric history. Comparison is made between those whose recent diagnosis was in the past year (original finding) vs in the past three years.

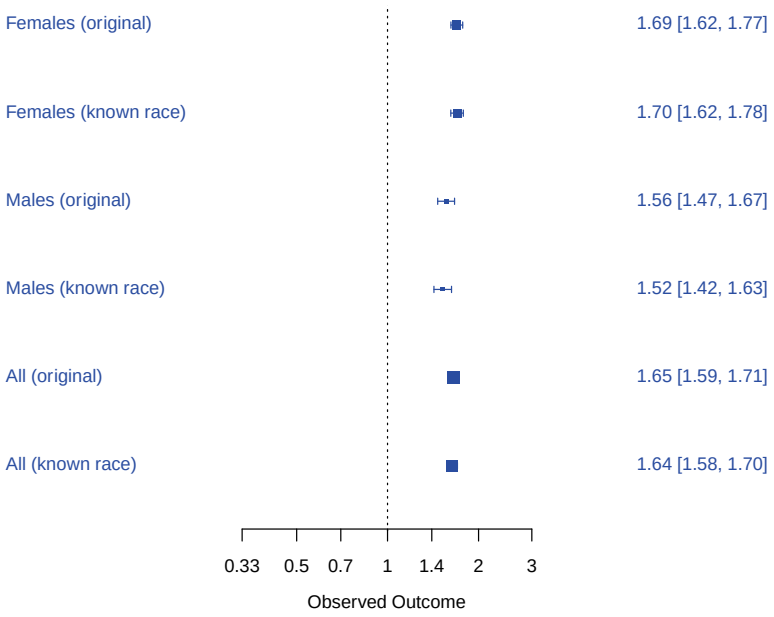

Fig. 18 – Relative risk of developing COVID-19 among patients with (vs without) a recent psychiatric history. Comparison is made between the relative risk among patients with known race vs all patients (original analysis).

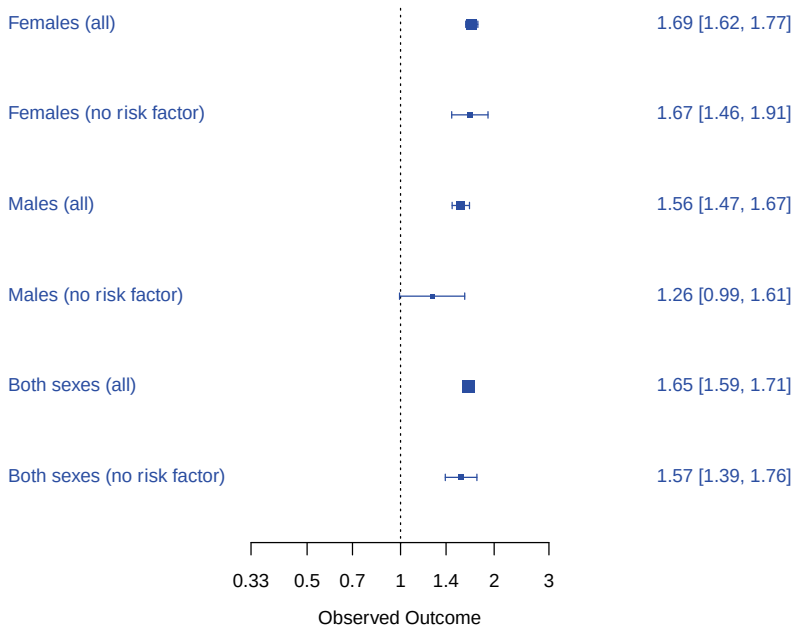

Fig. 19 – Relative risk of developing COVID-19 among patients with (vs without) a recent psychiatric history. Comparison is made between the relative risk among patients who have none of the physical risk factors for COVID-19 vs all patients (original analysis).

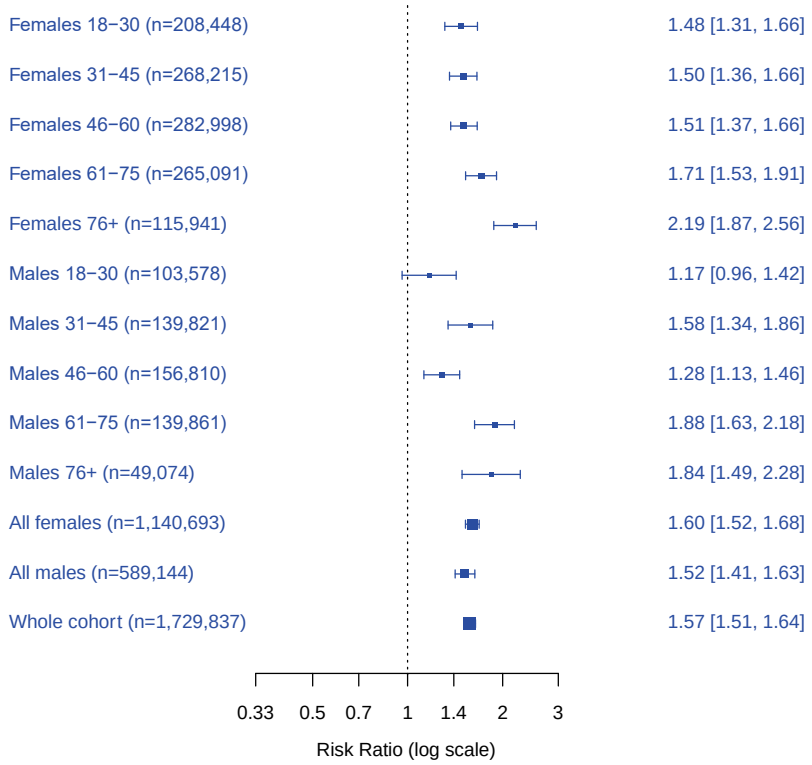

Fig. 20 - Relative risks of *confirmed* COVID-19 among patients with a psychiatric illness recorded in the past year compared to a matched cohort of patients with no history of psychiatric illness.

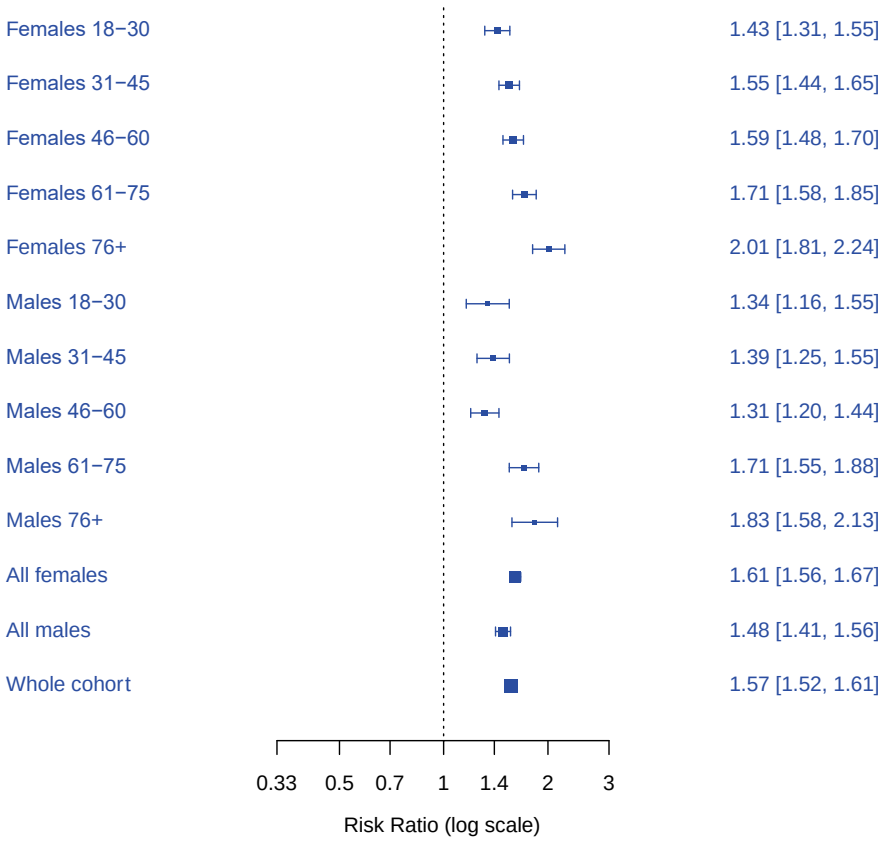

Fig. 21 - Relative risks of COVID-19 among patients with a psychiatric illness recorded in the past year compared to a matched cohort of patients with no history of psychiatric illness, wherein socioeconomic factors encoded using the Z59 code ('Problems related to housing and economic circumstances') were used in the matching process.

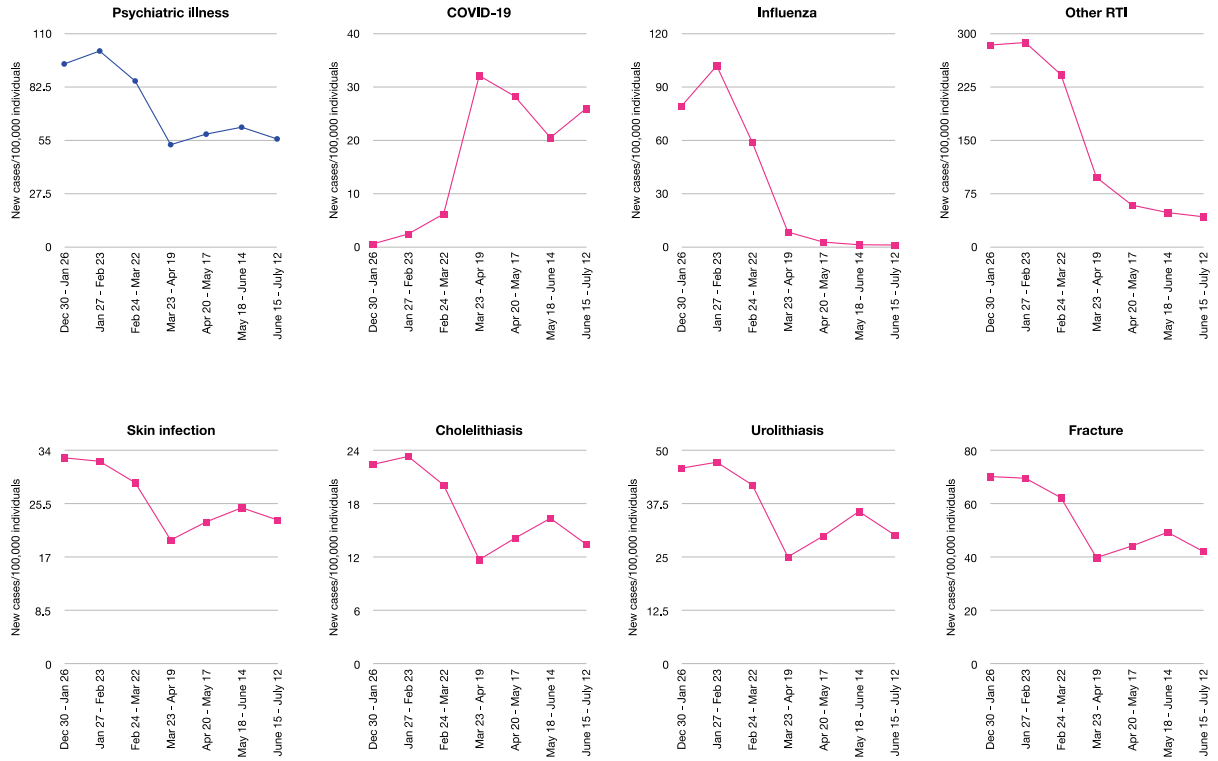

Fig. 22 – Incidence of first psychiatric diagnoses, COVID-19, and other health events during the study period. In contrast to COVID-19, all other health events had a decreased incidence from February to April.

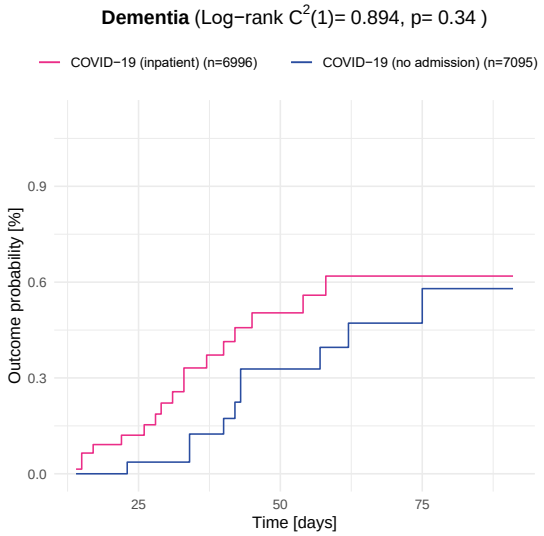

Fig. 23 – Kaplan-Meier curves comparing the incidence of dementia in the 14-90 days after COVID-19 between patients requiring and not requiring inpatient admission. The absence of a significant difference suggests that the increased rate of dementia diagnosis after COVID-19 is not merely a result of misdiagnosis of delirium for the latter is more common among inpatients (as is observed in our dataset; data not shown). Shaded areas represent 95% confidence intervals.

References

1 Kahn MG, Callahan TJ, Barnard J, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. EGEMS (Wash DC) 2016; 4: 1244.

2 Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999; **10**: 37–48.

3 Williamson EJ, Walker AJ, Bhaskaran K, *et al.* OpenSAFELY: factors associated with COVID-19 death in 17 million patients. *Nature* 2020; published online July 8. DOI:10.1038/s41586-020-2521-4.

4 Wiersinga WJ, Joost Wiersinga W, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19). JAMA. 2020; 324: 782.

5 https://www.cdc.gov/coronavirus/2019-ncov/community/strategy-discontinue-isolation.html (last visited: August 16, 2020)

6 Raisi-Estabragh Z, McCracken C, Bethell MS, et al. Greater risk of severe COVID-19 in Black, Asian and Minority Ethnic populations is not explained by cardiometabolic, socioeconomic or behavioural factors, or by 25 (OH)-vitamin D status: study of 1326 cases from the UK Biobank. J Public Health 2020; 42: 451–60.

7 Austin PC. Type I error rates, coverage of confidence intervals, and variance estimation in propensity-score matched analyses. Int J Biostat 2009; 5: Article 13.

8 Schemper M, Wakounig S, Heinze G. The estimation of average hazard ratios by weighted Cox regression. *Stat Med* 2009; **28**: 2473–89.

9 Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Statistics in Medicine. 2002; **21**: 2175–97.

10 Hastie T, Tibshirani R. Varying-coefficient models. *J R Stat Soc Series B Stat Methodol* 1993; **55**: 757–79.

11 Therneau T, Crowson C, Atkinson E. Using time dependent covariates and time dependent coefficients in the cox model. *Survival Vignettes* 2017. https://cran.pau.edu.tr/web/packages/survival/vignettes/timedep.pdf.

RECORD statement

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	Title and Abstract	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Abstract (Methods) Abstract (Methods) N/A
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	Abstract, Research in Context, and Introduction		
Objectives	3	State specific objectives, including any prespecified hypotheses	Abstract (Background) and Introduction		
Methods					
Study Design	4	Present key elements of study design early in the paper	Abstract and Methods (first subsection)		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Methods ('Data and study design', 'Analysis of psychiatric sequelae', and 'Analysis of psychiatric antecedents'		
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed	Methods ('Data and study design', 'Analysis of psychiatric sequelae', and 'Analysis of psychiatric antecedents' Methods ('Data and study design', 'Analysis of psychiatric sequelae', 'Analysis of psychiatric	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	Methods and appendix Appendix ('TriNetX network – Quality control') N/A

		Case-control study - For matched studies, give matching criteria and the number of controls per case	antecedents', 'Statistical analysis', and appendix Tables 1–7.		
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	Methods ('Variables of interest', 'Analysis of psychiatric sequelae, 'Analysis of psychiatric antecedents') and appendix	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Methods ('Variables of interest', 'Analysis of psychiatric sequelae, 'Analysis of psychiatric antecedents') and appendix
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Methods ('Variables of interest', 'Analysis of psychiatric sequelae, 'Analysis of psychiatric antecedents'), Discussion, and appendix		
Bias	9	Describe any efforts to address potential sources of bias	Methods ('Analysis of psychiatric sequelae, 'Analysis of psychiatric antecedents') and appendix		
Study size	10	Explain how the study size was arrived at	Methods ('Analysis of psychiatric sequelae, 'Analysis of psychiatric antecedents')		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	Methods ('Variables of interest and their coding') and appendix.		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	Methods ('Analysis of psychiatric sequelae, 'Analysis of psychiatric antecedents', and 'Statistical analysis')		
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	Author contribution section Appendix

Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	N/A
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	Results ('Characteristics of the cohorts and descriptive statistics') and Table 1 and appendix Tables 1–7.	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.*, study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	Methods section and Table 1 and appendix Tables 1–7.
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)	(a) Table 1 (b) Table 1 (c) Fig. 1 and Fig. 2		
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures	Fig. 1, Fig. 2, Fig. 3, Table 2, Table 3, Results section, appendix		
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Fig. 1, Fig. 2, Fig. 3, Table 2, Table 3, Results section, appendix N/A N/A		
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	Results section and appendix		
Discussion					
Key results	18	Summarise key results with reference to study objectives	Discussion first paragraph		

Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Discussion	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Discussion
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Discussion and Abstract		
Generalisability	21	Discuss the generalisability (external validity) of the study results	Discussion		
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Declaration of interest		
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Data sharing section

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.